University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2016 *Paper* 349

Data-adaptive Inference of the Optimal Treatment Rule and its Mean Reward. The Masked Bandit

Antoine Chambaz* Wenjing Zheng†

Mark J. van der Laan‡

*Université Paris Ouest Nanterre, achambaz@u-paris10.fr

†University of California, Berkeley, Division of Biostatistics, wenjing.zheng@berkeley.edu

‡University of California, Berkeley, Division of Biostatistics, laan@berkeley.edu

Data-adaptive Inference of the Optimal Treatment Rule and its Mean Reward. The Masked Bandit

Antoine Chambaz, Wenjing Zheng, and Mark J. van der Laan

Abstract

This article studies the data-adaptive inference of an optimal treatment rule. A treatment rule is an individualized treatment strategy in which treatment assignment for a patient is based on her measured baseline covariates. Eventually, a reward is measured on the patient. We also infer the mean reward under the optimal treatment rule. We do so in the so called non-exceptional case, i.e., assuming that there is no stratum of the baseline covariates where treatment is neither beneficial nor harmful, and under a companion margin assumption.

Our pivotal estimator, whose definition hinges on the targeted minimum loss estimation (TMLE) principle, actually infers the mean reward under the current estimate of the optimal treatment rule. This data-adaptive statistical parameter is worthy of interest on its own. Our main result is a central limit theorem which enables the construction of confidence intervals on both mean rewards under the current estimate of the optimal treatment rule and under the optimal treatment rule itself. The asymptotic variance of the estimator takes the form of the variance of an efficient influence curve at a limiting distribution, allowing to discuss the efficiency of inference.

As a by product, we also derive confidence intervals on two cumulated pseudo-regrets, a key notion in the study of bandits problems. Seen as two additional data-adaptive statistical parameters, they compare the sum of the rewards actually received during the course of the experiment with, either the sum of the means of the rewards, or the counterfactual rewards we would have obtained if we had used from the start the current estimate of the optimal treatment rule to assign treatment.

A simulation study illustrates the procedure. One of the cornerstones of the theoretical study is a new maximal inequality for martingales with respect to the uniform entropy integral.

1 Introduction

This article studies the data-adaptive inference of an optimal treatment rule. A treatment rule is an individualized treatment strategy in which treatment assignment for a patient is based on her measured baseline covariates. Eventually, a reward is measured on the patient. We also infer the mean reward under the optimal treatment rule.

The authors of [4] present an excellent unified overview on the estimation of optimal treatment rules, with a special interest in dynamic rules (where treatment assignment consists in successive assignments at successive time points). The estimation of the optimal treatment rule from independent and identically distributed (i.i.d.) observations has been studied extensively, with a recent interest in the use of machine learning algorithms to reach this goal [20, 31, 32, 29, 30, 23, 17]. Here, we estimate the optimal treatment rule (and its mean reward) based on sequentially sampled dependent observations by empirical risk minimization over sample-size-dependent classes of candidate estimates with a complexity controlled in terms of uniform entropy integral.

The estimation of the mean reward under the optimal treatment rule is more challenging than that of the optimal treatment rule. In [31, 32], the theoretical risk bound evaluating the statistical performance of the estimator of the optimal treatment rule can also be interpreted in terms of a measure of statistical performance of the resulting estimator of the mean reward under the optimal treatment rule. However, it does not yield confidence intervals.

Constructing confidence intervals for the mean reward under the optimal treatment rule is known to be more difficult when there exists a stratum of the baseline covariates where treatment is neither beneficial nor harmful [21]. In this so called "exceptional" case, the definition of the optimal treatment rule has to be disambiguated. Assuming non-exceptionality, confidence intervals are obtained in [29] for the mean reward under the (sub-) optimal treatment rule defined as the optimal treatment rule over a parametric class of candidate treatment rules, and in [15] for the actual mean reward under the optimal treatment rule. In the more general case where exceptionality can occur, different approaches have been considered [5, 11, 14, 16]. Here, we focus on the non-exceptional case under a companion margin assumption [18].

We rely on the targeted minimum loss estimation (TMLE) principle [26, 25]. We can build upon previous studies on the construction and statistical analysis of targeted, covariate-adjusted, response-adaptive trials also based on TMLE [6, 33, 7]. One of the cornerstones of the theoretical study is a new maximal inequality for martingales with respect to (wrt) the uniform entropy integral, proved by decoupling [8], symmetrization and chaining, which allows us to control several empirical processes indexed by random functions.

Our pivotal TMLE estimator is actually constructed as an estimator of the mean reward under the current estimate of the optimal treatment rule. Worthy of interest on its own, this data-adaptive statistical parameter (or similar ones) has also been considered in [5, 13, 14, 15, 16]. Our main result is a central limit theorem for our TMLE estimator. The asymptotic variance takes the form of the variance of an efficient influence curve at a limiting distribution, allowing to discuss the efficiency of inference.

We use our TMLE estimator to infer the mean rewards under the current estimate of the optimal treatment rule and under the optimal treatment rule itself. Moreover, we use it to infer two additional data-adaptive statistical parameters. The first one compares the sum of the rewards actually received during the course of the experiment with the sum of the means of the rewards we would have obtained if we had used from the start the current estimate of the optimal treatment rule to assign treatment. The second one compares the sum of the rewards actually received during the course of the experiment with the sum of the counterfactual rewards we would have obtained if we had used from the start the current estimate of the optimal treatment rule to assign treatment.

Both additional data-adaptive statistical parameters are "cumulated pseudo-regrets". We borrow this expression from the literature on bandits. Bandits have raised a considerable interest in the machine learning community as relevant models for interactive learning schemes or recommender systems. Many articles define efficient strategies to minimize the expected cumulated pseudo-regret (also known as the "cumulated regret"), see [3] for a survey. Sometimes, the objective is to identify the arm with the largest mean reward (the best arm) as fast and accurately as possible, regardless of the number of times a sub-optimal arm is played, see [10] for an in-depth analysis of the so called fixed-confidence setting where one looks for a strategy guaranteeing that the probability of wrongly identifying the best arm at some stopping time is no more than a fixed maximal risk while minimizing the stopping time's expectation. Here, we derive confidence intervals on the cumulated pseudo-regrets as by products of the confidence intervals that we build for the mean rewards under the current estimate of the optimal treatment rule and under the optimal treatment rule itself. Thus, the most relevant comparison is with the so called "contextual bandit problems", see [12, Chapter 4] for an excellent overview.

Organization

Section 2 presents our targeted, data-adaptive sampling scheme and our pivotal estimator. Section 3 studies the convergence of the sampling scheme, *i.e.*, how the sequences of stochastic and treatment rules converge, assuming that a function of the conditional mean of the reward given treatment and baseline covariate is consistently estimated. Section 4 is devoted to the presentation of our main result, a central limit theorem for our pivotal estimator, to the comment of its assumptions and to an example. Section 5 builds upon the previous section to build confidence intervals for the mean rewards under the current estimate of the optimal treatment rule and under the optimal treatment rule itself, as well as confidence intervals for the two cumulated pseudo-regrets evoked in the introduction. Section 6 presents the results of a simulation study. Section 7 closes the article with a brief discussion. All proofs are given in Appendix A. Technical lemmas are gathered in Appendix B and C.

2 Targeting the optimal treatment rule and its mean reward

2.1 Statistical setting

At sample size n, we will have observed the ordered vector $\mathbf{O}_n \equiv (O_1, \ldots, O_n)$, with convention $O_0 \equiv \emptyset$. For every $1 \leq i \leq n$, the data structure O_i writes as $O_i \equiv (W_i, A_i, Y_i)$. Here, $W_i \in \mathcal{W}$ consists of the baseline covariates (some of which may be continuous) of the ith patient, $A_i \in \mathcal{A} \equiv \{0, 1\}$ is the binary treatment of interest assigned to her, and $Y_i \in \mathcal{Y}$ is her primary outcome of interest. We interpret Y as a *reward*: the larger is Y, the better. We assume that the reward space $\mathcal{O} \equiv \mathcal{W} \times \mathcal{A} \times \mathcal{Y}$ is bounded. Without loss of generality, we may then assume that $\mathcal{Y} \equiv (0, 1)$, *i.e.*, that the rewards are between and bounded away from 0 and 1. Interestingly, the content of this article would still hold up to minor modifications if we assumed instead $\mathcal{Y} \equiv \{0, 1\}$.

Let μ_W be a measure on $\mathcal{W}$ equipped with a σ-field, $\mu_A = \text{Dirac}(0) + \text{Dirac}(1)$ be a measure on $\mathcal{A}$ equipped with its σ-field, and μ_Y be the Lebesgue measure on $\mathcal{Y}$ equipped with the Borel σ-field. Define $\mu \equiv \mu_W \otimes \mu_A \otimes \mu_Y$, a measure on $\mathcal{O}$ equipped with the product of the above σ-fields. The unknown, true likelihood of $\mathbf{O}_n$ wrt $\mu^{\otimes n}$ is given by the following factorization of the density of $\mathbf{O}_n$ wrt $\mu^{\otimes n}$:

$$\begin{aligned} \mathcal{L}_{Q_0,\mathbf{g}_n}(\mathbf{O}_n) &\equiv \prod_{i=1}^{n} Q_{W,0}(W_i) \times (A_i g_i(1|W_i) + (1 - A_i) g_i(0|W_i)) \times Q_{Y,0}(Y_i|A_i, W_i) \\ &= \prod_{i=1}^{n} Q_{W,0}(W_i) \times g_i(A_i|W_i) \times Q_{Y,0}(Y_i|A_i, W_i), \end{aligned} \tag{1}$$

where *(i)* $w \mapsto Q_{W,0}(w)$ is the density wrt μ_W of a true, unknown law on $\mathcal{W}$ (that we assume being dominated by μ_W), *(ii)* $\{y \mapsto Q_{Y,0}(y|a, w) : (a, w) \in \mathcal{A} \times \mathcal{W}\}$ is the collection of the conditional densities $y \mapsto Q_{Y,0}(y|a, w)$ wrt μ_Y of true, unknown laws on $\mathcal{Y}$ indexed by (a, w) (that we assume being all dominated by μ_Y), *(iii)* $g_i(1|W_i)$ is the known conditional probability that $A_i = 1$ given W_i, and *(iv)* $\mathbf{g}_n \equiv (g_1, \ldots, g_n)$, the ordered vector of the n first *stochastic rules*. One reads in (1) *(i)* that $W_1, \ldots, W_n$ are independently sampled from $Q_{W,0} d\mu_W$, *(ii)* that $Y_1, \ldots, Y_n$ are conditionally sampled from $Q_{Y,0}(\cdot|A_1, W_1) d\mu_Y, \ldots, Q_{Y,0}(\cdot|A_n, W_n) d\mu_Y$, respectively, and *(iii)* that each A_i is drawn conditionally on W_i from the Bernoulli distribution with known parameter $g_i(1|W_i)$.

We introduce the semiparametric collection $\mathcal{Q}$ of all elements of the form

$$\begin{aligned} Q &= (Q_W d\mu_W, Q_Y(\cdot|a, w), (a, w) \in \mathcal{A} \times \mathcal{W}), \quad \text{or} \\ Q &= \left(Q_W \sum_{k=1}^{K} \text{Dirac}(w_k), Q_Y(\cdot|a, w), (a, w) \in \mathcal{A} \times \mathcal{W} \right) \end{aligned}$$

with $\{w_1, \ldots, w_K\} \subset \mathcal{W}$. Here, Q_W is a density wrt either μ_W or a discrete measure $\sum_{k=1}^{K} \text{Dirac}(w_k)$ (thus, we can take the empirical measure of W as first component of Q). Each $Q_Y(\cdot|a, w)$ is a density wrt μ_Y. In particular, $Q_0 \equiv (Q_{W,0} d\mu_W, Q_{Y,0}(\cdot|a, w), (a, w) \in \mathcal{A} \times \mathcal{W}) \in \mathcal{Q}$. In light of (1) define, for every $Q \in \mathcal{Q}$, $\mathcal{L}_{Q,\mathbf{g}_n}(\mathbf{O}_n) \equiv \prod_{i=1}^{n} Q_W(W_i) \times g_i(A_i|W_i) \times Q_Y(Y_i|A_i, W_i)$. The set $\{\mathcal{L}_{Q,\mathbf{g}_n} : Q \in \mathcal{Q}\}$ is a semiparametric model for the likelihood of $\mathbf{O}_n$. It contains the true, unknown likelihood $\mathcal{L}_{Q_0,\mathbf{g}_n}$.

Fix arbitrarily $Q \in \mathcal{Q}$. The conditional expectation of Y given (A, W) under Q is denoted $Q_Y(A, W) \equiv \int y Q_Y(y|A, W) d\mu_Y(y)$. To alleviate notation, we introduce the so called "blip function" q_Y characterized by $q_Y(W) = Q_Y(1, W) - Q_Y(0, W)$. If $q_Y(W) \geq 0$ (respectively, $q_Y(W) < 0$), then assigning treatment $A = 1$ (respectively, $A = 0$) guarantees that the patient receives the superior treatment in the sense that her mean reward is larger in this arm than in the other one. If $q_Y(W) = 0$, then the mean rewards are equal. This characterizes an optimal stochastic rule $r(Q_Y)$ given by

$$r(Q_Y)(W) \equiv \mathbf{1}\{q_Y(W) \geq 0\}. \tag{2}$$

It is degenerate because, given W, the assignment is deterministic. Such degenerate stochastic rules are usually referred to as *treatment rules* in the causal inference literature. When $Q = Q_0$, we denote $Q_Y \equiv Q_{Y,0}$, $q_Y \equiv q_{Y,0}$, and $r(Q_Y) \equiv r_0$.

The parameter of interest is the mean reward under the optimal treatment rule,

$$\psi_0 \equiv E_{Q_0}\left(Q_{Y,0}(r_0(W), W)\right) = \int Q_{Y,0}(r_0(w), w) Q_{W,0}(w) d\mu_W(w).$$

Let $\mathcal{G}$ be the semiparametric collection of all stochastic treatment rules g, which satisfy $g(1|W) = 1 - g(0|W) \in (0,1)$. From now on, for each $(Q, g) \in \mathcal{Q} \times \mathcal{G}$, we denote $P_{Q,g}$ the distribution of $O = (W, A, Y)$ obtained by drawing W from Q_W, then A from the Bernoulli distribution with parameter $g(1|W)$, then Y from the conditional distribution $Q_Y(\cdot|A, W)d\mu_Y$. Let $\mathcal{M} \equiv \{P_{Q,g} : Q \in \mathcal{Q}, g \in \mathcal{G}\}$. We actually see ψ_0 as the value at any $P_{Q_0,g}$ $(g \in \mathcal{G})$ of the mapping $\Psi : \mathcal{M} \to [0,1]$ characterized by

$$\Psi(P_{Q,g}) \equiv E_Q\left(Q_Y(r(Q_Y)(W), W)\right).$$

Obviously, the parameter $\Psi(P_{Q,g})$ does not depend on g. It depends linearly on the marginal distribution $Q_W d\mu_W$, but in a more subtle way on the conditional expectation Q_Y.

We have not specified yet what is precisely $\mathbf{g}_n \equiv (g_1, \ldots, g_n)$. Our targeted sampling scheme "targets" the optimal treatment rule r_0 and ψ_0. By targeting r_0, we mean estimating $Q_{Y,0}$ based on past observations, and relying on the resulting estimator to collect the next block of data, as seen in (1), and to estimate ψ_0. Targeting ψ_0 refers to our efforts to build an estimator of ψ_0 which allows the construction of valid, narrow confidence intervals.

2.2 Targeted, data-adaptive sampling and inference

Let $\{t_n\}_{n\geq 1}$ and $\{\xi_n\}_{n\geq 1}$ be two user-supplied, non-increasing sequences with $t_1 \leq 1/2$, $\lim_n t_n \equiv t_\infty > 0$ and $\lim_n \xi_n \equiv \xi_\infty > 0$. For every $n \geq 1$, introduce the function G_n characterized over $[-1, 1]$ by

$$\begin{aligned} G_n(x) &= t_n \mathbf{1}\{x \leq -\xi_n\} \\ &\quad + \left(-\frac{1/2 - t_n}{2\xi_n^3} x^3 + \frac{1/2 - t_n}{2\xi_n/3} x + \frac{1}{2}\right) \mathbf{1}\{-\xi_n \leq x \leq \xi_n\} \\ &\quad + (1 - t_n) \mathbf{1}\{x \geq \xi_n\}. \end{aligned}$$

For convenience, we also introduce $G_\infty \equiv G_{n_1}$ where $n_1 \geq 1$ is chosen large enough so that $t_{n_1} = t_\infty$ and $\xi_{n_1} = \xi_\infty$. Function G_n is non-decreasing and c_n-Lipschitz with

$$c_n \equiv \frac{1/2 - t_n}{2\xi_n/3} \leq \frac{1/2 - t_\infty}{2\xi_\infty/3} \equiv c_\infty.$$

This particular choice of G_n is one among many. Any other non-decreasing function $\tilde{G}_n$ such that $\tilde{G}_n(x) = t_n$ for $x \leq -\xi_n$, $\tilde{G}_n(x) = 1 - t_n$ for $x \geq \xi_n$, and $\tilde{G}_n$ κ_n-Lipschitz with κ_n upper-bounded by a finite κ_∞ could be chosen as well.

Loss functions and working models. Let $g^{\mathrm{b}} \in \mathcal{G}$ be the balanced stochastic rule wherein each arm is assigned with probability $1/2$ regardless of baseline covariates. Let $g^{\mathrm{ref}} \in \mathcal{G}$ be a stochastic rule, bounded away from 0 and 1 by choice, that serves as a reference. In addition, let L be a loss function for $Q_{Y,0}$ and $\mathcal{Q}_{1,n}$ be a working model

$$\mathcal{Q}_{1,n} \equiv \{Q_{Y,\beta} : \beta \in B_n\} \subset \mathcal{Q}_Y \equiv \{Q_Y : Q \in \mathcal{Q}\}$$

consisting of functions $Q_{Y,\beta}$ mapping $\mathcal{A} \times \mathcal{W}$ to $[0,1]$ (in the above display, Q_Y denotes the conditional expectation of Y given (A,W) under $Q \in \mathcal{Q}$). One choice of L is the quasi negative-log-likelihood loss function L^{kl}. For any $Q_Y \in \mathcal{Q}_Y$ bounded away from 0 and 1, $L^{\mathrm{kl}}(Q_Y)$ satisfies

$$-L^{\mathrm{kl}}(Q_Y)(O) \equiv Y \log(Q_Y(A,W)) + (1-Y)\log\left(1 - Q_Y(A,W)\right).$$

Another interesting loss function L for $Q_{Y,0}$ is the least-square loss function L^{ls}. It is characterized at any $Q_Y \in \mathcal{Q}_Y$ by

$$L^{\mathrm{ls}}(Q_Y)(O) \equiv (Y - Q_Y(A,W))^2.$$

Completing the description of the sampling scheme. We initialize the sampling scheme by setting $g_1 \equiv g^{\mathrm{b}}$. Consider $1 < i \leq n$. Since

$$Q_{Y,0} = \underset{Q_Y \in \mathcal{Q}_Y}{\arg\min}\, E_{Q_0,g}(L(Q_Y)(O)),$$

we naturally define

$$\beta_i \in \underset{\beta \in B_i}{\arg\min}\, \frac{1}{i-1} \sum_{j=1}^{i-1} L(Q_{Y,\beta})(O_j) \frac{g^{\mathrm{ref}}(A_j|W_j)}{g_j(A_j|W_j)} \tag{3}$$

and use Q_{Y,β_i} as an estimator of $Q_{Y,0}$ based on $\mathbf{O}_{i-1}$. It gives rise to q_{Y,β_i} and r_i such that

$$\begin{aligned} q_{Y,\beta_i}(W) &\equiv Q_{Y,\beta_i}(1,W) - Q_{Y,\beta_i}(0,W), \\ r_i(W) &\equiv \mathbf{1}\{q_{Y,\beta_i}(W) \geq 0\}, \end{aligned} \tag{4}$$

two substitution estimators of the blip function $q_{Y,0}$ and optimal treatment rule r_0, respectively.

For smaller sample sizes i, setting g_i equal to r_i would be hazardous. Indeed, there is no guarantee that q_{Y,β_i} estimates well $q_{Y,0}$. Say, for instance, that $q_{Y,\beta_i}(w)$ is large by mere chance for all $w \in D_i \subset \mathcal{W}$. If we used $g_i = r_i$, then future patients with $W \in D_i$ would systematically be assigned to treatment arm $a = 1$ and the poor estimation of $q_{Y,0}$ on D_i could not be corrected, if needed. Thus, we characterize g_i by setting

$$g_i(1|W) \equiv G_i(q_{Y,\beta_i}(W)).$$

This completes the definition of the likelihood function, hence the characterization of our sampling scheme.

Note that choosing $t_1 = \ldots = t_{n_0} = 1/2$ for a limit sample size n_0 would yield $g_1 = \ldots = g_{n_0} = g^{\mathrm{b}}$, the balanced stochastic rule. Furthermore, the definitions of G_n and g_n entail straightforwardly the following lemma:

Lemma 1. *Set $n \geq 1$. It holds that*

$$\inf_{w \in \mathcal{W}} g_n(r_n(w)|w) \geq 1/2, \tag{5}$$

$$\inf_{w \in \mathcal{W}} g_n(1 - r_n(w)|w) \geq t_n. \tag{6}$$

Lemma 1 illustrates the so called exploration/exploitation trade-off, *i.e.*, the ability of the sampling scheme to exploit the accrued information (5) while keeping exploring in search of potential discordant new piece of information (6). From a different perspective, (5) shows that treatment rule r_n meets the positivity assumption.

Targeted minimum loss estimator. Let $\mathcal{R}$ be the set of all treatment rules, *i.e.*, the set of all functions mapping $\mathcal{W}$ to $\{0,1\}$. For each $g \in \mathcal{G}$ and $\rho \in \mathcal{R}$, we define a function $H_\rho(g)$ mapping $\mathcal{O}$ to $\mathbb{R}$ by setting

$$H_\rho(g)(O) \equiv \frac{\mathbf{1}\{A = \rho(W)\}}{g(A|W)}. \tag{7}$$

Introduce the following one-dimensional parametric model for $Q_{Y,0}$:

$$\{Q_{Y,\beta_n,g_n,r_n}(\epsilon) \equiv \text{expit}\left(\text{logit}(Q_{Y,\beta_n}) + \epsilon H_{r_n}(g_n)\right) : \varepsilon \in \mathcal{E}\}, \tag{8}$$

where $\mathcal{E} \subset \mathbb{R}$ is a closed, bounded interval containing 0 in its interior. We characterize an optimal fluctuation parameter by setting

$$\epsilon_n \in \arg\min_{\epsilon \in \mathcal{E}} \frac{1}{n} \sum_{i=1}^{n} L^{\text{kl}}(Q_{Y,\beta_n,g_n,r_n}(\epsilon))(O_i) \frac{g_n(A_i|W_i)}{g_i(A_i|W_i)}. \tag{9}$$

Define $Q^*_{Y,\beta_n,g_n,r_n} \equiv Q_{Y,\beta_n,g_n,r_n}(\epsilon_n)$ and

$$\psi_n^* \equiv \frac{1}{n} \sum_{i=1}^{n} Q^*_{Y,\beta_n,g_n,r_n}(r_n(W_i), W_i). \tag{10}$$

Funded on the TMLE principle, ψ_n^* is our pivotal estimator.

3 Convergence

For every $p \geq 1$ and measurable $f : \mathcal{W} \to \mathbb{R}$, let $\|f\|_p$ be the seminorm given by

$$\|f\|_p^p \equiv \int |q_{Y,0}| \times |f|^p Q_{W,0} d\mu_W.$$

We introduce $g_0 \in \mathcal{G}$ given by

$$g_0(1|W) \equiv G_\infty(q_{Y,0}(W)). \tag{11}$$

The stochastic rule g_0 approximates the treatment rule r_0 in the following sense:

$$|g_0(1|W) - r_0(W)| \leq t_\infty \mathbf{1}\{|q_{Y,0}(W)| \geq \xi_\infty\} + \tfrac{1}{2}\mathbf{1}\{|q_{Y,0}(W)| < \xi_\infty\}. \tag{12}$$

Therefore, if t_∞ is small and if $|q_{Y,0}(W)| \geq \xi_\infty$, then drawing A from g_0 does not differ much from drawing A from r_0. Rigorously, the distance in total variation between the Bernoulli laws with parameters $g_0(1|W)$ and $r_0(W)$ equals $2t_\infty$. On the contrary, if $|q_{Y,0}(W)| < \xi_\infty$, then the conditional laws of A given W under g_0 or r_0 may be very different. However, if ξ_∞ is small, then assigning randomly $A = 1$ or $A = 0$ has little impact on the mean value of the reward Y.

We now study the convergence of r_n to r_0 and that of g_n to g_0. In each case, the convergence is relative to two measures of discrepancy. For r_n, we consider the seminorm $\|r_n - r_0\|_p$ (any $p \geq 1$) and

$$\Delta(r_n, r_0) \equiv |E_{Q_0,r_n}(Q_{Y,0}(A,W)) - E_{Q_0,r_0}(Q_{Y,0}(A,W))|. \tag{13}$$

By analogy, the measures of discrepancy for g_n are

$$\|g_n - g_0\|_p \equiv \|g_n(1|\cdot) - g_0(1|\cdot)\|_p, \quad (14)$$
$$\Delta(g_n, g_0) \equiv |E_{Q_0,g_n}(Q_{Y,0}(A,W)) - E_{Q_0,g_0}(Q_{Y,0}(A,W))|. \quad (15)$$

Note that $\Delta(r_n, r_0)$ and $\Delta(g_n, g_0)$ are the absolute values of the differences between the mean rewards under the treatment rules r_n and r_0 and the stochastic rules g_n and g_0, respectively. As such, they are targeted toward our end result, *i.e.*, the inference of ψ_0, as shown in the following lemma:

Lemma 2. *Set $n \geq 1$. It holds that*

$$0 \leq \psi_0 - E_{Q_0,r_n}(Q_{Y,0}(A,W)) = \Delta(r_n, r_0) \leq \|r_n - r_0\|_1, \quad (16)$$
$$0 \leq \psi_0 - E_{Q_0,g_n}(Q_{Y,0}(A,W)) \leq \Delta(g_n, g_0) + t_\infty + \xi_\infty. \quad (17)$$

The next lemma shows that the convergence of q_{Y,β_n} to $q_{Y,0}$ implies that of r_n to r_0.

Lemma 3. *Set $p \geq 1$. If $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1)$, then $\|r_n - r_0\|_p = o_P(1)$ hence $\Delta(r_n, r_0) = o_P(1)$.*

Similarly, the convergence of q_{Y,β_n} to $q_{Y,0}$ implies that of g_n to g_0.

Lemma 4. *Set $p \geq 1$. It holds that $0 \leq \Delta(g_n, g_0) \leq \|g_n - g_0\|_p$. Moreover, if $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1)$, then $\|g_n - g_0\|_p = o_P(1)$ hence $\Delta(g_n, g_0) = o_P(1)$.*

4 Asymptotia

4.1 Notation

Consider a class $\mathcal{F}$ of functions mapping a measured space $\mathcal{X}$ to $\mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}$. Recall that $\mathcal{F}$ is said separable if there exists a countable collection $\mathcal{F}'$ of functions such that each element of $\mathcal{F}$ is the pointwise limit of a sequence of elements of $\mathcal{F}'$. If $\phi \circ f$ is well defined for each $f \in \mathcal{F}$, then we note $\phi(\mathcal{F}) \equiv \{\phi \circ f : f \in \mathcal{F}\}$. In particular, we introduce the sets $\mathcal{G}_{1,n} \equiv \{G_n(q_Y) : Q_Y \in \mathcal{Q}_{1,n}\}$, $r(\mathcal{Q}_{1,n}) \equiv \{r(Q_Y) : Q_Y \in \mathcal{Q}_{1,n}\}$ (all $n \geq 1$) and $\mathcal{G}_1 \equiv \cup_{n \geq 1} \mathcal{G}_{1,n}$.

Set $\delta > 0$, μ a probability measure on $\mathcal{X}$, and let F be an envelope function for $\mathcal{F}$, *i.e.*, a function such that $|f(x)| \leq F(x)$ for every $f \in \mathcal{F}$, $x \in \mathcal{X}$. We denote $N(\delta, \mathcal{F}, \|\cdot\|_{2,\mu})$ the δ-covering number of $\mathcal{F}$ wrt $\|\cdot\|_{2,\mu}$, *i.e.*, the minimum number of $L^2(\mu)$-balls of radius δ needed to cover $\mathcal{F}$. The corresponding uniform entropy integral wrt F for $\mathcal{F}$ evaluated at δ is $J_F(\delta, \mathcal{F}) \equiv \int_0^\delta \sqrt{\log \sup_\mu N(\varepsilon \|F\|_{2,\mu}, \mathcal{F}, \|\cdot\|_{2,\mu})} d\varepsilon$, where the supremum is taken over all probability measures μ on the measured space $\mathcal{X}$ for which $\|F\|_{2,\mu} > 0$.

In general, given a known $g \in \mathcal{G}$ and an observation O drawn from $P_{Q_0,g}$, $Z \equiv g(A|W)$ is a deterministic function of g and O. Note that Z should be interpreted as a weight associated with O and will be used as such. Therefore, we can augment O with Z, *i.e.*, substitute (O, Z) for O, while still denoting $(O, Z) \sim P_{Q_0,g}$. In particular, during the course of our trial, conditionally on $\mathbf{O}_{i-1}$, the stochastic rule g_i is known and we can substitute $(O_i, Z_i) = (O_i, g_i(A_i|W_i)) \sim P_{Q_0,g_i}$ for O_i drawn from P_{Q_0,g_i}. The inverse weights $1/g_i(A_i|W_i)$ are bounded because $\mathcal{G}_1$ is uniformly bounded away from 0 and 1.

The empirical distribution of $\mathbf{O}_n$ is denoted P_n. For a measurable function $f : \mathcal{O} \times [0,1] \to \mathbb{R}^d$, we use the notation $P_n f \equiv n^{-1} \sum_{i=1}^n f(O_i, Z_i)$. Likewise, for any fixed $P_{Q,g} \in \mathcal{M}$, $P_{Q,g} f \equiv E_{Q,g}(f(O,Z))$ and, for each $i = 1, \ldots, n$,

$$\begin{aligned} P_{Q_0,g_i} f &\equiv E_{Q_0,g_i}[f(O_i, Z_i)|\mathbf{O}_{i-1}], \\ P_{Q_0,\mathbf{g}_n} f &\equiv \frac{1}{n} \sum_{i=1}^n E_{Q_0,g_i}[f(O_i, Z_i)|\mathbf{O}_{i-1}]. \end{aligned}$$

The supremum norm of a function $f : \mathcal{O} \times [0,1] \to \mathbb{R}^d$ is denoted $\|f\|_\infty$. When $d = 1$, we denote $\|f\|^2_{2,P_{Q_0,g^{\text{ref}}}} \equiv P_{Q_0,g^{\text{ref}}} f^2$. If f is only a function of W, then $\|f\|_2 = \big\| |q_{Y,0}|^{1/2} f \big\|_{2,P_{Q_0,g^{\text{ref}}}}$.

For every $Q_{Y,\beta} \in \mathcal{Q}_1 \equiv \cup_{n \geq 1} \mathcal{Q}_{1,n}$, the blip function $Q_{Y,\beta}(1,\cdot) - Q_{Y,\beta}(0,\cdot)$ is denoted $q_{Y,\beta}$ by analogy with (4). We will often deal with seminorms $\|f\|_2$ with $f = Q_Y - Q_{Y,\beta_0}$ for some $Q_Y \in \mathcal{Q}_Y$ and $\mathcal{Q}_{Y,\beta_0} \in \mathcal{Q}_1$. A consequence of the trivial inequality $(a-b)^2 \leq 2(ua^2 + (1-u)b^2)/\min(u, 1-u)$ (valid for all $a, b \in \mathbb{R}$, $0 < u < 1$), the following bound will prove useful:

$$\begin{aligned} \|q_Y - q_{Y,\beta_0}\|_2 &\leq 2 \left\| |q_{Y,0}|^{1/2} / g^{\text{ref}} \right\|_\infty \times \|Q_Y - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} \\ &\leq 2 \|1/g^{\text{ref}}\|_\infty \times \|Q_Y - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}}. \end{aligned} \tag{18}$$

The constant $2\|1/g^{\text{ref}}\|_\infty$ is minimized at $g^{\text{ref}} = g^{\text{b}}$, with $2\|1/g^{\text{b}}\|_\infty = 4$.

4.2 Central limit theorem

Our main result is a central limit theorem for ψ_n^*. It relies on the following assumptions, upon which we comment in Section 4.3.

A1. The conditional distribution of Y given (A, W) under Q_0 is not degenerated. Moreover, $P_{Q_0}(|q_{Y,0}(W)| > 0) = 1$.

Existence and convergence of projections.

A2. For each $n \geq 1$, there exists $Q_{Y,\beta_{n,0}} \in \mathcal{Q}_{1,n}$ satisfying

$$P_{Q_0,g^{\text{ref}}} L(Q_{Y,\beta_{n,0}}) = \inf_{Q_{Y,\beta} \in \mathcal{Q}_{1,n}} P_{Q_0,g^{\text{ref}}} L(Q_{Y,\beta}).$$

Moreover, there exists $Q_{Y,\beta_0} \in \mathcal{Q}_1$ such that, for all $\delta > 0$,

$$P_{Q_0,g^{\text{ref}}} L(Q_{Y,\beta_0}) < \inf_{\left\{Q_Y \in \mathcal{Q}_1 : \|Q_Y - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} \geq \delta\right\}} P_{Q_0,g^{\text{ref}}} L(Q_Y).$$

Finally, it holds that $q_{Y,\beta_0} = q_{Y,0}$.

A3. For all $\rho \in \mathcal{R}$ and $\epsilon \in \mathcal{E}$, introduce

$$Q_{Y,\beta_0,g_0,\rho}(\epsilon) \equiv \text{expit}\left(\text{logit}(Q_{Y,\beta_0}) + \epsilon H_\rho(g_0)\right), \tag{19}$$

where $H_\rho(g_0)$ is given by (7) with $g = g_0$. For every $\rho \in \mathcal{R}$, there exists a unique $\epsilon_0(\rho) \in \mathcal{E}$ such that

$$\epsilon_0(\rho) \in \arg\min_{\epsilon \in \mathcal{E}} P_{Q_0,g_0} L^{\text{kl}}(Q_{Y,\beta_0,g_0,\rho}(\epsilon)). \tag{20}$$

Reasoned complexity.

A4. The classes $\mathcal{Q}_{1,n}$, $L(\mathcal{Q}_{1,n})$ and $r(\mathcal{Q}_{1,n})$ are separable. Moreover, the following entropy conditions hold: $J_1(1, \mathcal{Q}_{1,n}) = o(\sqrt{n})$, $J_1(1, r(\mathcal{Q}_{1,n})) = o(\sqrt{n})$, $J_{F_n}(1, L(\mathcal{Q}_{1,n})) = o(\sqrt{n})$, where each F_n is an envelope function for $L(\mathcal{Q}_{1,n})$.

A4*. Let $\{\delta_n\}_{n\geq 1}$ be a sequence of positive numbers. If $\delta_n = o(1)$, then $J_1(\delta_n, \mathcal{Q}_{1,n}) = o(1)$ and $J_1(\delta_n, r(\mathcal{Q}_{1,n})) = o(1)$.

Margin condition.

A5. There exist $\gamma_1, \gamma_2 > 0$ such that, for all $t \geq 0$,

$$P_{Q_0}\left(0 < |q_{Y,0}(W)| \leq t\right) \leq \gamma_1 t^{\gamma_2}.$$

We first focus on the convergence of the sequences of stochastic rules g_n and empirical treatment rule r_n. By Lemmas 3 and 4, it suffices to consider the convergence of q_{Y,β_n}. By (18), we may consider the convergence of Q_{Y,β_n}.

Proposition 1. *Under* **A2** *and* **A4**, *both* $\|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ *and* $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1)$. *Hence, for any* $p \geq 1$, $\|r_n - r_0\|_p = o_P(1)$, $\|g_n - g_0\|_p = o_P(1)$, $\Delta(r_n, r_0) = o_P(1)$, $\Delta(g_n, g_0) = o_P(1)$ *by Lemmas 3 and 4. If* **A1** *and* **A5** *are also met, then* $\|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ *and* $\|g_n - g_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ *as well.*

Define now the data-adaptive parameter

$$\psi_{r_n,0} \equiv E_{Q_0}\left(Q_{Y,0}(r_n(W), W)\right) = E_{Q_0,r_n}\left(Q_{Y,0}(A, W)\right). \tag{21}$$

By (16) in Lemma 2 and Lemma 3, we have the following corollary to Proposition 1:

Corollary 1. *Under* **A2** *and* **A4**, $0 \leq \psi_0 - \psi_{r_n,0} = o_P(1)$.

We now turn to the convergence of ψ_n^*. Its asymptotic behavior can be summarized in these terms:

Theorem 1. *Suppose that* **A1**, **A2**, **A3**, **A4**, **A4*** *and* **A5** *are met. It holds that* $\psi_n^* - \psi_{r_n,0} = o_P(1)$. *Thus, by Corollary 1,* $\psi_n^* - \psi_0 = o_P(1)$ *as well. Moreover,* $\sqrt{n/\Sigma_n}(\psi_n^* - \psi_{r_n,0})$ *is approximately standard normally distributed, where* Σ_n *is the explicit estimator given in (30).*

Theorem 1 is a toned down version of Theorem 2 that we state and comment on in Section 4.5. Section 4.3 discusses their assumptions and Section 4.4 presents an example. Theorems 1 and 2 allow the construction of confidence intervals for several parameters of interest, as shown in Section 5.

4.3 Commenting on the assumptions

Assumption **A1** consists in two statements. The first one is a simple condition guaranteeing that the limit variance of $\sqrt{n}(\psi_n^* - \psi_{r_n,0})$ is positive. The second one is more stringent. In the terminology of [21], it states that Q_0 is not exceptional. If Q_0 were exceptional, then the set $\{w \in \mathcal{W} : q_{Y,0}(w) = 0\}$ would have positive probability under Q_0. To a patient falling in this set, the optimal treatment rule $r(Q_{Y,0}) \equiv r_0$ recommends to assign treatment $A = 1$ instead of treatment $A = 0$. This arbitrary choice has no consequence whatsoever in terms of conditional mean of the reward given treatment and baseline covariates.

However, it is well documented that exceptional laws are problematic. For the estimation of the optimal treatment rule r_0, one reason is that an estimator will typically not converge to a fixed limit on $\{w \in \mathcal{W} : q_{Y,0}(w) = 0\}$ [21, 22, 16]. Another reason is that the mean reward under the optimal treatment rule seen as a functional, Ψ, is pathwise differentiable at Q_0 if and only if, Q_0-almost surely, either $|q_{Y,0}(W)| > 0$ or the conditional distributions of Y given $(A = 1, W)$ and $(A = 0, W)$ under Q_0 are degenerated [16, Theorem 1]. This explains why it is also assumed that the true law is not exceptional in [29, 15, 17]. Other approaches have been considered to circumvent the need to make this assumption: relying on m-out-of-n bootstrap [5] (at the cost of a $\sqrt{m} = o(\sqrt{n})$-rate of convergence and need to fine-tune m), or changing the parameter of interest by focusing on the mean reward under the optimal treatment rule conditional on patients for whom the best treatment has a clinically meaningful effect (truncation) [11, 13, 14].

To the best of our knowledge, only [16] addresses the inference of the original parameter at a $\sqrt{n}$-rate of convergence without assuming that the true law is not exceptional. Moreover, if the true law is not exceptional, then the estimator is asymptotically efficient among all regular and asymptotically linear estimators. Developed in the i.i.d. setting, the estimator of [16] does not require that the estimator of r_0 converge as the sample size grows. It relies on a clever iteration of a two-step procedure consisting in *(i)* estimating well-chosen nuisance parameters, including r_0, on a small chunk of data, then *(ii)* constructing an estimator targeted to the mean reward under the current estimate of r_0 with the nuisance parameters obtained in *(i)*. The final estimator is a weighted average of the resulting chunk-specific estimators. Adapting this procedure to our setting where data are dependent would be very challenging.

Assumptions **A2** states the existence of L-projections $Q_{Y,\beta_{n,0}}$ of $Q_{Y,0}$ onto each working model $\mathcal{Q}_{1,n}$ and their convergence to a limit L-projection $Q_{Y,\beta_0} \in \mathcal{Q}_1 \equiv \cup_{n\geq 1}\mathcal{Q}_{1,n}$. More importantly, it states that the blip function q_{Y,β_0} associated with Q_{Y,β_0} equals the true blip function $q_{Y,0}$ associated with $Q_{Y,0}$.

For any fixed treatment rule $\rho \in \mathcal{R}$, the limit L-projection Q_{Y,β_0} can be fluctuated in a direction $H_\rho(g_0)$ characterized by ρ and $Q_{Y,0}$, see (7), (11)) and (19). Assumption **A3** states that there exists a unique L^{kl}-projection of $Q_{Y,0}$ onto this ρ-specific one-dimensional parametric model fluctuating Q_{Y,β_0}. In particular, when $\rho = r_n$, the estimator of r_0 at sample size n, $Q_{Y,0}$ is uniquely L^{kl}-projected onto, say, $Q^*_{Y,0,r_n}$. One of the keys to our approach is the equality $E_{Q_0}(Q^*_{Y,0,r_n}(r_n(W), W)) = \psi_{r_n,0} \equiv E_{Q_0}(Q_{Y,0}(r_n(W), W))$ even if $Q_{Y,0}$ and $Q^*_{Y,0,r_n}$ differ. Proven in step 3 of the proof of Proposition 7, which states that ψ_n^* is a consistent estimator of $\psi_{r_n,0}$ (*i.e.*, $\psi_n^* - \psi_{r_n,0} = o_P(1)$), this robustness property is a by product of the robustness of the efficient influence curve of the mean reward under r_n treated as a fixed treatment rule, see Lemma 12.

Expressed in terms of separability and conditions on uniform entropy integrals, **A4** and **A4*** restrict the complexities of the working models $\mathcal{Q}_{1,n}$ and resulting classes $r(\mathcal{Q}_{1,n})$ and $L(\mathcal{Q}_{1,n})$. Imposing separability is a convenient way to ensure that some delicate measurability conditions are met. Assumption **A4*** partially strengthens **A4** because choosing $\delta_n \equiv 1/\sqrt{n}$ (all $n \geq 1$) in **A4*** implies $J_1(1, \mathcal{F}_n) = o(\sqrt{n})$ for both $\mathcal{F}_n \equiv \mathcal{Q}_{1,n}$ and $\mathcal{F}_n \equiv r(\mathcal{Q}_{1,n})$ by a simple change of variable. Section 4.4 presents an example of sequence

$\{\mathcal{Q}_{1,n}\}_{n\geq 1}$ of working models which meets **A4** and **A4***. Its construction involves VC-classes of functions, which are archetypical examples of classes with well-behaved uniform entropy integrals. Restricting the complexities of the working models $\mathcal{Q}_{1,n}$, $r(\mathcal{Q}_{1,n})$ and $L(\mathcal{Q}_{1,n})$ in terms of bracketing entropy is tempting because of the great diversity of examples of classes of functions which behave well in these terms. Unfortunately, this is not a viable alternative, since bounds on the bracketing numbers of $\mathcal{Q}_{1,n}$ do not imply bounds on those of $r(\mathcal{Q}_{1,n})$.

Inspired from the seminal article [18], assumptions similar to **A5** are known as "margin assumptions" in the literature. They describe how the data-distribution concentrates on adverse events, *i.e.*, on events that make inference more difficult. We have already discussed the fact that inferring the optimal treatment rule and its mean reward is less challenging when the law of the absolute value of $|q_{Y,0}(W)|$ puts no mass on $\{0\}$. It actually occurs that the less mass this law puts *around* $\{0\}$, the less challenging is the inference. Assumption **A5** formalizes tractable concentrations. It has already proven useful in the i.i.d. setting, see [15, Lemma 1] and [16, Condition (16)]. By Markov's inequality, **A5** is implied by the following, clearer assumption:

A5.** There exists $\gamma_2 > 0$ such that

$$\gamma_1 \equiv E_{Q_0}\left(|q_{Y,0}(W)|^{-\gamma_2}\mathbf{1}\{|q_{Y,0}(W)| > 0\}\right) < \infty.$$

4.4 An example

In this section, we construct a sequence $\{\mathcal{Q}_{1,n}\}_{n\geq 1}$ of working models which meets **A4** and **A4***, see Proposition 2. Let $\mathcal{F}^-$ be a separable class of measurable functions from $\mathcal{W}$ to $[-1,1]\setminus\{0\}$ such that $\{\{w \in \mathcal{W} : f^-(w) \geq t\} : f^- \in \mathcal{F}^-, t \in [-1,1]\}$ is a VC-class of sets. By definition, $\mathcal{F}^-$ is a VC-major class [28, Sections 2.6.1 and 2.6.4]. Thus, Corollary 2.6.12 in [28] guarantees the existence of two constants $K^- > 0$ and $\alpha^- \in [0,1)$ such that, for every $\varepsilon > 0$,

$$\log \sup_{\mu} N(\varepsilon\|1\|_{2,\mu}, \mathcal{F}^-, \|\cdot\|_{2,\mu}) \leq K^- \left(\frac{1}{\varepsilon}\right)^{2\alpha^-}. \tag{22}$$

Let $\mathcal{F}^+$ be a separable class of measurable functions from $\mathcal{W}$ to $[0,2]$ such that, for two constants $K^+ > 0$, $\alpha^+ \in [0,1)$ and for every $\varepsilon > 0$,

$$\log \sup_{\mu} N(\varepsilon\|2\|_{2,\mu}, \mathcal{F}^+, \|\cdot\|_{2,\mu}) \leq K^+ \left(\frac{1}{\varepsilon}\right)^{2\alpha^+}. \tag{23}$$

For instance, $\mathcal{F}^+$ may be a VC-hull class of functions, *i.e.*, a subset of the pointwise sequential closure of the symmetric convex hull of a VC-class of functions [28, Section 2.6.3]. (The suprema in (22) and (23) are taken over all probability measures μ on the measured space $\mathcal{W}$.)

We now use $\mathcal{F}^-$ and $\mathcal{F}^+$ to define the sequence $\{\mathcal{Q}_{1,n}\}_{n\geq 1}$ of working models. Let $\mathcal{F}^- = \cup_{n\geq 1}\mathcal{F}_n^-$ and $\mathcal{F}^+ = \cup_{n\geq 1}\mathcal{F}_n^+$ be rewritten as the limits of two increasing sequences of sets $\{\mathcal{F}_n^-\}_{n\geq 1}$ and $\{\mathcal{F}_n^+\}_{n\geq 1}$. Set $n \geq 1$ and define

$$B_n \equiv \{(f^-, f^+) \in \mathcal{F}_n^- \times \mathcal{F}_n^+ : 0 \leq f^+ + f^-, f^+ - f^- \leq 2\}.$$

For each $\beta \equiv (f^-, f^+) \in B_n$, introduce $Q_{Y,\beta}$ mapping $\mathcal{A} \times \mathcal{W}$ to $[0,1]$ characterized by

$$Q_{Y,\beta}(A, W) = \frac{A}{2}(f^+(W) + f^-(W)) + \frac{(1-A)}{2}(f^+(W) - f^-(W)). \tag{24}$$

We define the nth working model as $\mathcal{Q}_{1,n} \equiv \{Q_{Y,\beta} : \beta \in B_n\}$. It is separable because $\mathcal{F}^-$ and $\mathcal{F}^+$ are separable.

Because $q_{Y,\beta} \equiv Q_{Y,\beta}(1,\cdot) - Q_{Y,\beta}(0,\cdot) = f^-$ for every $\beta \equiv (f^-, f^+) \in B_n$, it holds that

$$
\begin{aligned}
r(\mathcal{Q}_{1,n}) &\equiv \{\mathbf{1}\{q_{Y,\beta}(\cdot) \geq 0\} : \beta \in B_n\} \\
&= \left\{\mathbf{1}\{f^-(\cdot) \geq 0\} : f^- \in \mathcal{F}_n^-\right\} \subset \left\{\mathbf{1}\{f^-(\cdot) \geq 0\} : f^- \in \mathcal{F}^-\right\}
\end{aligned}
$$

which, by construction, is a fixed subset of a VC-class of functions, hence a VC-class of functions itself. Moreover, $r(\mathcal{Q}_{1,n})$ is separable because $\mathcal{F}^-$ is separable and elements of $\mathcal{F}^-$ take only positive or negative values. These properties and (22), (23) are the main arguments in the proof of the following result:

Proposition 2. *The sequence $\{\mathcal{Q}_{1,n}\}_{n\geq 1}$ of working models satisfies* **A4** *(with $L = L^{ls}$ the least-square loss) and* **A4***.

4.5 Asymptotic linear expansion and resulting central limit theorem

Theorem 1 is a summary of Theorem 2 below, whose main result is the asymptotic linear expansion (31). The statement of Theorem 2 requires additional notation.

Let $Q_{Y,0}^*$, $d_{W,0}^*$, $d_{Y,0}^*$ and Σ_0 be given by

$$
\begin{aligned}
Q_{Y,0}^*(A,W) &\equiv Q_{Y,\beta_0,g_0,r_0}(\epsilon_0(r_0))(A,W), \\
d_{W,0}^*(W) &\equiv Q_{Y,0}^*(r_0(W),W) - E_{Q_0}(Q_{Y,0}^*(r_0(W),W)), && (25)\\
d_{Y,0}^*(O,Z) &\equiv \frac{\mathbf{1}\{A = r_0(W)\}}{Z}(Y - Q_{Y,0}^*(A,W)), && (26)\\
\Sigma_0 &\equiv P_{Q_0,g_0}(d_{W,0}^* + d_{Y,0}^*)^2. && (27)
\end{aligned}
$$

Analogously, recall that $Q_{Y,\beta_n,g_n,r_n}^* \equiv Q_{Y,\beta_n,g_n,r_n}(\epsilon_n)$ and let $d_{W,n}^*$, $d_{Y,n}^*$ and Σ_n be given by

$$
\begin{aligned}
d_{W,n}^*(W) &\equiv Q_{Y,\beta_n,g_n,r_n}^*(r_n(W),W) - \psi_n^*, && (28)\\
d_{Y,n}^*(O,Z) &\equiv \frac{\mathbf{1}\{A = r_n(W)\}}{Z}(Y - Q_{Y,\beta_n,g_n,r_n}^*(A,W)), && (29)\\
\Sigma_n &\equiv P_n(d_{W,n}^* + d_{Y,n}^*)^2. && (30)
\end{aligned}
$$

Note that $d_{W,n}^*$, $d_{Y,n}^*$ and Σ_n are empirical counterparts to $d_{W,0}^*$, $d_{Y,0}^*$ and Σ_0.

Theorem 2. *Suppose that* **A1**, **A2**, **A3**, **A4**, **A4*** *and* **A5** *are met. It holds that $\psi_n^* - \psi_{r_n,0} = o_P(1)$. Thus, by Corollary 1, $\psi_n^* - \psi_0 = o_P(1)$ as well. Moreover, $\Sigma_n = \Sigma_0 + o_P(1)$ with $\Sigma_0 > 0$ and*

$$
\psi_n^* - \psi_{r_n,0} = (P_n - P_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + d_{W,0}^*) + o_P(1/\sqrt{n}). \quad (31)
$$

Consequently, $\sqrt{n/\Sigma_n}(\psi_n^* - \psi_{r_n,0})$ converges in law to the standard normal distribution.

Consider (25). It actually holds that the centering term $E_{Q_0}(Q_{Y,0}^*(r_0(W),W))$ equals $\psi_0 \equiv E_{Q_0}(Q_{Y,0}(r_0(W),W))$ (see step one of the proof of Corollary 2 in Section A.2). This proximity between $Q_{Y,0}$ and $Q_{Y,0}^*$ follows from the careful fluctuation of Q_{Y,β_0}.

Set $Q_0^* \equiv (Q_{W,0} d\mu_W, Q_{Y,0}^*(\cdot|a,w), (a,w) \in \mathcal{A} \times \mathcal{W}) \in \mathcal{Q}$. The influence function $d_{Y,0}^* + d_{W,0}^*$ in (31) is closely related to the efficient influence curve $D_{r_0}(Q_0^*, g_0)$ at $P_{Q_0^*, g_0}$ of the mapping $\Psi_{r_0} : \mathcal{M} \to [0,1]$ characterized by

$$\Psi_{r_0}(P_{Q,g}) \equiv E_Q \left(Q_Y(r_0(W), W) \right),$$

the mean reward under Q of the treatment rule r_0 (possibly different from the optimal treatment rule $r(Q_Y)$ under Q) treated as known and fixed. Specifically, in light of Lemma 12 in Section C,

$$d_{Y,0}^*(O,Z) + d_{W,0}^*(W) = D_{r_0}(Q_0^*, g_0)(O)$$

when $Z = g_0(A|W)$. Consequently, $\Sigma_0 = P_{Q_0,g_0} D_{r_0}(Q_0^*, g_0)^2$.

If $Q_{Y,\beta_0} = Q_{Y,0}$ (a stronger condition than equality $q_{Y,\beta_0} = q_{Y,0}$ in **A2**), then $Q_{Y,0}^* = Q_{Y,0}$ (because $\epsilon_0(r_0)$ from **A3** equals zero) hence $Q_0^* = Q_0$ and, finally, the remarkable equality $\Sigma_0 = P_{Q_0,g_0} D_{r_0}(Q_0, g_0)^2$: the asymptotic variance of $\sqrt{n}(\psi_n^* - \psi_{r_n,0})$ coincides with the generalized Cramér-Rao lower bound for the asymptotic variance of any regular and asymptotically linear estimator of $\Psi_{r_0}(P_{Q_0,g_0})$ when sampling independently from P_{Q_0,g_0} (see Lemma 12). Otherwise, the discrepancy between Σ_0 and $P_{Q_0,g_0} D_{r_0}(Q_0, g_0)^2$ will vary depending on that between Q_{Y,β_0} and $Q_{Y,0}$, hence in particular on the user-supplied sequence $\{\mathcal{Q}_{1,n}\}_{n \geq 1}$ of working models. Studying this issue in depth is very difficult, if at all possible, and beyond the scope of this article.

5 Confidence regions

We explore how Theorems 1 and 2 enable the construction of confidence intervals for various possibly data-adaptive parameters: the mean rewards under the optimal treatment rule and under its current estimate in Section 5.1; the empirical cumulative pseudo-regret in Section 5.2; the counterfactual cumulative pseudo-regret in Section 5.3.

Set a confidence level $\alpha \in (0, 1/2)$. Let $\xi_\alpha < 0$ and $\xi_{1-\alpha/2} > 0$ be the corresponding α- and $(1-\alpha/2)$-quantiles of the standard normal distribution.

5.1 Confidence intervals for the mean rewards under the optimal treatment rule and under its current estimate

Theorems 1 and 2 yield straightforwardly a confidence interval for the mean reward under the current best estimate of the optimal treatment rule, $\psi_{r_n,0}$.

Proposition 3. *Under the assumptions of Theorems 1 or 2, the probability of the event*

$$\psi_{r_n,0} \in \left[\psi_n^* \pm \xi_{1-\alpha/2} \sqrt{\frac{\Sigma_n}{n}} \right]$$

converges to $(1-\alpha)$ *as* n *goes to infinity.*

We need to strengthen **A5** to guarantee that the confidence interval in Proposition 3 can also be used to infer the mean reward under the optimal treatment rule, ψ_0. Consider thus the following.

A5*. There exist $\gamma_1 > 0$, $\gamma_2 \geq 1$ such that, for all $t \geq 0$,

$$P_{Q_0}\left(0 < |q_{Y,0}(W)| \leq t\right) \leq \gamma_1 t^{\gamma_2}.$$

Just like **A5** is a consequence of **A5****, **A5*** is a consequence of **A5**** where one substitutes the condition $\gamma_2 > 0$ for the stronger condition $\gamma_2 \geq 1$.

Proposition 4. *Under* **A5**** *there exists a constant $c > 0$ such that*

$$0 \leq \psi_0 - \psi_{r_n,0} \leq c\|q_{Y,\beta_n} - q_{Y,0}\|_2^{2(1+\gamma_2)/(3+\gamma_2)}. \tag{32}$$

Set $\gamma_3 \equiv 1/4 + 1/2(1+\gamma_2) \in (1/4, 1/2]$. By (32), if $\|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1/n^{\gamma_3})$, then $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1/n^{\gamma_3})$, which implies $0 \leq \psi_0 - \psi_{r_n,0} = o_P(1/\sqrt{n})$.

Therefore, if the assumptions of Theorems 1 or 2 are also met, then the probability of the event

$$\psi_0 \in \left[\psi_n^* \pm \xi_{1-\alpha/2}\sqrt{\frac{\Sigma_n}{n}}\right]$$

converges to $(1-\alpha)$ as n goes to infinity.

The definition of γ_3 in Proposition 4 justifies the requirement $\gamma_2 \geq 1$ in **A5***. Indeed, $\gamma_3 \leq 1/2$ is equivalent to $\gamma_2 \geq 1$. Moreover, it holds that $\gamma_3 = 1/2$ (so that $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1/n^{\gamma_3})$ can be read as a parametric rate of convergence) if and only if $\gamma_2 = 1$.

5.2 Lower confidence bound for the empirical cumulative pseudo-regret

We call

$$\mathcal{E}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(Y_i - Q_{Y,0}(r_n(W_i), W_i)) \tag{33}$$

the "empirical cumulative pseudo-regret" at sample size n. A data-adaptive parameter, it is the difference between the average of the *actual* rewards garnered so far, $n^{-1}\sum_{i=1}^{n} Y_i$, and the average of the *mean* rewards under the current estimate r_n of the optimal treatment rule r_0 in the successive contexts drawn so far during the course of the experiment, $n^{-1}\sum_{i=1}^{n} Q_{Y,0}(r_n(W_i), W_i)$. The former is a known quantity, so the real challenge is to infer the latter. Moreover, we are mainly interested in obtaining a lower confidence bound.

Define

$$\begin{aligned}
\Sigma_0^{\mathcal{E}} &\equiv E_{Q_0,g_0}\left(d_{W,0}^*(W) - (Q_{Y,0}(r_0(W), W) - \psi_0) + d_{Y,0}^*(O,Z)\right)^2, \\
\Sigma_n^{\mathcal{E}} &\equiv \frac{1}{n}\sum_{i=1}^{n}\left(d_{W,n}^*(W_i) - \left(Q_{Y,\beta_n}(r_n(W_i), W_i) - \psi_n^0\right) + d_{Y,n}^*(O_i, Z_i)\right)^2,
\end{aligned}$$

with $\psi_n^0 \equiv n^{-1}\sum_{i=1}^{n} Q_{Y,\beta_n}(r_n(W_i), W_i)$. Note that $\Sigma_n^{\mathcal{E}}$ is an empirical counterpart to $\Sigma_0^{\mathcal{E}}$.

Proposition 5. *Under the assumptions of Theorems 1 or 2, the probability of the event*

$$\mathcal{E}_n \geq \frac{1}{n}\sum_{i=1}^{n} Y_i - \psi_n^* + \xi_\alpha\sqrt{\frac{\Sigma_n^{\mathcal{E}}}{n}}$$

converges to $(1-\alpha)$ as n goes to infinity.

5.3 Lower confidence bound for the counterfactual cumulative pseudo-regret

In this section, we cast our probabilistic model in a causal model. We postulate the existence of counterfactual rewards $Y_n(1)$ and $Y_n(0)$ of assigning treatment $a = 1$ and $a = 0$ to the nth patient (all $n \geq 1$). They are said counterfactual because it is impossible to observe them jointly. The observed nth reward writes $Y_n = A_n Y_n(1) + (1 - A_n)Y_n(0)$.

We call

$$\mathcal{C}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i(r_n(W_i))) \tag{34}$$

the "counterfactual cumulative pseudo-regret" at sample size n. It is the difference between the average of the *actual* rewards garnered so far, $n^{-1}\sum_{i=1}^{n} Y_i$, and the average of the *counterfactual* rewards under the current estimate r_n of the optimal treatment rule r_0 in the successive contexts drawn so far during the course of the experiment, $n^{-1}\sum_{i=1}^{n} Y_i(r_n(W_i))$. Once more, the former is a known quantity, so the real challenge is to infer the latter. Moreover, we are mainly interested in obtaining a lower confidence bound.

For simplicity, we adopt the so called "non-parametric structural equations" approach [19]. So, we actually postulate the existence of a sequence $\{U_n\}_{n\geq 1}$ of i.i.d. random variables independent from $\{O_n\}_{n\geq 1}$ with values in $\mathcal{U}$ and that of a deterministic measurable function $\mathbb{Q}_{Y,0}$ mapping $\mathcal{A} \times \mathcal{W} \times \mathcal{U}$ to $\mathcal{Y}$ such that, for every $n \geq 1$ and both $a = 0, 1$,

$$Y_n(a) = \mathbb{Q}_{Y,0}(a, W_n, U_n).$$

The notation $\mathbb{Q}_{Y,0}$ is motivated by the following property. Let $(A, W, U) \in \mathcal{A} \times \mathcal{W} \times \mathcal{U}$ be distributed from $\mathbb{P}$ in such a way that *(i)* A is conditionally independent from U given W, and *(ii)* with $Y \equiv A\mathbb{Q}_{Y,0}(1, W, U) + (1 - A)\mathbb{Q}_{Y,0}(0, W, U)$, the conditional distribution of Y given (A, W) is $Q_{Y,0}(\cdot|A, W)d\mu_Y$. Then, for each $a \in \mathcal{A}$,

$$\begin{aligned} E_{\mathbb{P}}\left(\mathbb{Q}_{Y,0}(a, W, U)|W\right) &= E_{\mathbb{P}}\left(\mathbb{Q}_{Y,0}(a, W, U)|A = a, W\right) \\ &= E_{\mathbb{P}}\left(Y|A = a, W\right) \\ &= Q_{Y,0}(a, W). \end{aligned} \tag{35}$$

Although $\mathcal{C}_n$ is by nature a counterfactual data-adaptive parameter, it is possible to construct a conservative lower confidence bound yielding a confidence interval whose asymptotic coverage is no less than $(1 - \alpha)$.

Proposition 6. *Under the assumptions of Theorems 1 or 2, the probability of the event*

$$\mathcal{C}_n \geq \frac{1}{n}\sum_{i=1}^{n} Y_i - \psi_n^* + \xi_\alpha \sqrt{\frac{\Sigma_n^{\mathcal{E}}}{n}}$$

converges to $(1 - \alpha') \geq (1 - \alpha)$ *as* n *goes to infinity.*

The key to this result is threefold. First, the asymptotic linear expansion (31) still holds in the above causal model where each observation (O_n, Z_n) is augmented with U_n (every $n \geq 1$). Second, the expansion yields a confidence interval with asymptotic level $(1-\alpha)$. Unfortunately, its asymptotic width depends on features of the causal distribution which are not identifiable from the real-world (as opposed to causal) distribution. Third, and fortunately, $\Sigma_n^{\mathcal{E}}$ is a conservative estimator of the limit width. We refer the reader to the proof of Proposition 6 in Section A.3 for details. It draws inspiration from [1], where the same trick was first devised to estimate the so called sample average treatment effect.

Linear contextual bandit problems. Consider the following contextual bandit problem: an agent is sequentially presented a context $w_t \in \mathbb{R}^d$, has to choose an action $a_t \in \{0,1\}$, and receives a random reward $y_t = f(a_t, w_t) + \varepsilon_t$, with f an unknown real-valued function and ε_t a centered, typically sub-Gaussian noise. The agent aims at maximizing the cumulated sum of rewards. The contextual bandit problem is linear if there exists $\theta \equiv (\theta_0, \theta_1) \in \mathbb{R}^{2d}$ such that $f(a,w) \equiv w^\top \theta_a$ for all $(a,w) \in \{0,1\} \times \mathbb{R}^d$. At time t, the best action is $a_t^* \equiv \arg\max_{a=0,1} w_t^\top \theta_a$ and maximizing the cumulated sum of rewards is equivalent to minimizing the cumulated pseudo-regret $R_T^\theta \equiv \sum_{t=1}^T w_t^\top (a_t^* \theta_{a_t^*} - a_t \theta_{a_t})$.

We refer to [12, Chapter 4] for an overview of the literature dedicated to this problem, which bears evident similitudes with our problem of interest. Optimistic algorithms consist in constructing a frequentist region of confidence for θ and choosing that action a_{t+1} maximizing $a \mapsto \max_\vartheta w_{t+1}^\top \vartheta_a$ where ϑ ranges over the confidence region. The Bayes-UCB algorithm and its variants follow the same idea with Bayesian regions of confidence substituted for the frequentist ones. As for the celebrated Thompson Sampling algorithm, it consists in drawing $\tilde{\theta}$ from the posterior distribution of θ and choosing that action a_{t+1} maximizing $a \mapsto w_{t+1}^\top \tilde{\theta}_a$. Each time estimating θ (which is essentially equivalent to estimating the optimal treatment rule and its mean reward) is a means to an end.

Various frequentist analyses of such algorithms have been proposed. It notably appears that the cumulated pseudo-regret R_T^θ typically scales in $\tilde{O}(\sqrt{T})$ with high probability, where $\tilde{O}$ ignores logarithmic factors in T. This is consistent with the form of the lower confidence bounds that we obtain, as by products rather than main objectives and under milder assumptions on $f/Q_{Y,0}$, for our empirical and counterfactual cumulated pseudo-regrets.

6 Simulation study

6.1 Setup

We now present the results of a simulation study. Under Q_0, the baseline covariate W decomposes as $W \equiv (U, V) \in [0,1] \times \{1,2,3\}$, where U and V are independent random variables respectively drawn from the uniform distribution on $[0,1]$ and such that $P_{Q_0}(V = 1) = \frac{1}{2}$, $P_{Q_0}(V = 2) = \frac{1}{3}$ and $P_{Q_0}(V = 3) = \frac{1}{6}$. Moreover, Y is conditionally drawn given (A, W) from the Beta distribution with a constant variance set to 0.1 and a mean $Q_{Y,0}(A, W)$ satisfying

$$
\begin{aligned}
Q_{Y,0}(1,W) &\equiv \frac{1}{2}\left(1 + \frac{3}{4}\cos(\pi U V)\right), \\
Q_{Y,0}(0,W) &\equiv \frac{1}{2}\left(1 + \frac{1}{2}\sin(3\pi U / V)\right).
\end{aligned}
$$

The conditional means and associated blip function $q_{Y,0}$ are represented in Figure 2 (left plots). We compute the numerical values of the following parameters: $\psi_0 \approx 0.6827$ (true parameter); $\mathrm{Var}_{P_{Q_0,g^\mathrm{b}}} D(Q_0, g^\mathrm{b})(O) \approx 0.1916^2$ (the variance under P_{Q_0,g^b} of the efficient influence curve of Ψ at P_{Q_0,g^b}, *i.e.*, under Q_0 with equiprobability of being assigned $A = 1$ or $A = 0$); $\mathrm{Var}_{P_{Q_0,g_0}} D(Q_0, g_0)(O) \approx 0.1666^2$ (the variance under P_{Q_0,g_0} of the efficient influence curve of Ψ at P_{Q_0,g_0}, *i.e.*, under Q_0 and the approximation g_0 to the optimal treatment rule r_0); and $\mathrm{Var}_{P_{Q_0,r_0}} D(Q_0, r_0)(O) \approx 0.1634^2$ (the variance under P_{Q_0,r_0} of the efficient influence curve of Ψ at P_{Q_0,r_0}, *i.e.*, under Q_0 and the optimal treatment rule r_0).

The sequences $\{t_n\}_{n\geq1}$ and $\{\xi_n\}_{n\geq1}$ are chosen constant, with values $t_\infty = 10\%$ and $\xi_\infty = 1\%$ respectively. We choose $g^{\mathrm{ref}} = g^{\mathrm{b}}$ as reference. The targeting steps are performed when sample size is a multiple of 100, at least 200 and no more than 1000, when sampling is stopped. At such a sample size n, the working model $\mathcal{Q}_{1,n}$ consists of functions $Q_{Y,\beta}$ mapping $\mathcal{A} \times \mathcal{W}$ to $[0,1]$ such that, for each $a \in \mathcal{A}$ and $v \in \{1,2,3\}$, $\operatorname{logit} Q_{Y,\beta}(a,(U,v))$ is a linear combination of $1, U, U^2, \ldots, U^{d_n}$ and $\mathbf{1}\{(l-1)/\ell_n \leq U < l/\ell_n\}$ $(1 \leq l \leq \ell_n)$ with $d_n = 3 + \lfloor n/500 \rfloor$ and $\ell_n = \lceil n/250 \rceil$. The resulting global parameter β belongs to $\mathbb{R}^{6(d_n+\ell_n+1)}$ (in particular, $\mathbb{R}^{60}$ at sample size $n = 1000$). Working model $\mathcal{Q}_{1,n}$ is fitted wrt $L = L^{\mathrm{kl}}$ using the `cv.glmnet` function from package `glmnet` [9], with weights given in (3) and the option `"lambda.min"`. This means imposing (data-adaptive) upper-bounds on the ℓ^1- and ℓ^2-norms of parameter β (via penalization), hence the search for a sparse optimal parameter β_n.

6.2 Results

We repeat $N = 1000$ times, independently, the procedure described in Section 2.2 and the construction of confidence intervals for $\psi_{r_n,0}$ and confidence lower-bounds for the empirical and counterfactual cumulative pseudo-regrets described in Section 5. We report in Table 1 four empirical summary measures computed across simulations for each parameter among $\psi_{r_n,0}$, ψ_0, $\mathcal{E}_n$ and $\mathcal{C}_n$. In rows ^{a}: the empirical coverages. In rows ^{b} and ^{c}: the p-values of the binomial tests of 95%-coverage at least or 94%-coverage at least (null hypotheses) against their one-sided alternatives. In rows ^{d}: the mean values of the possibly data-adaptive parameters. In rows ^{e}: the mean values of Σ_n (for $\psi_{r_n,0}$), mean values of $|\mathcal{E}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{E}_n|$ (for $\mathcal{E}_n$), mean values of $|\mathcal{C}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{C}_n|$ (for $\mathcal{C}_n$).

It appears that the empirical coverage of the confidence intervals for the data-adaptive parameter $\psi_{r_n,0}$ and the fixed parameter ψ_0 is very satisfying. Although 14 out of 18 empirical proportions of coverage lie below 95%, the simulation study does not reveal a coverage smaller than 94%, even without adjusting for multiple testing. For sample size larger than 400, the simulation study does not reveal a coverage smaller than the nominal 95%, even without adjusting for multiple testing.

The asymptotic variance of ψ_n^* seems to stabilize below 0.1850^2. This is slightly smaller than $\operatorname{Var}_{P_{Q_0,g^{\mathrm{b}}}} D(Q_0,g^{\mathrm{b}})(O) \approx 0.1916^2$ $(1916/1850 \approx 1.04)$ and a little larger than $\operatorname{Var}_{P_{Q_0,g_0}} D(Q_0,g_0)(O) \approx 0.1666^2$ $(1850/1666 \approx 1.11)$. In theory, the asymptotic variance of ψ_n^* can converge to $\operatorname{Var}_{P_{Q_0,g_0}} D(Q_0,g_0)(O)$ if Q_{Y,β_n} converges to $Q_{Y,0}$. Rigorously speaking, this cannot be the case here given the working models we rely on. This is nonetheless a quite satisfying finding: we estimate $\psi_{r_n,0}$ and ψ_0 more efficiently than if we had achieved their efficient estimation based on i.i.d. data sampled under Q_0 and the balanced treatment rule g^{b} and, in addition, do so in such a way that most patients (those for whom $r_n(W) = r_0(W)$) are much more likely (90% versus 50%) to be assigned their respective optimal treatments.

The empirical coverage provided by the lower confidence bounds on the data-adaptive parameters $\mathcal{E}_n$ and $\mathcal{C}_n$ is excellent. Actually, the empirical proportions of coverage for $\mathcal{E}_n$, all larger than 96.5%, suggest that either $\mathcal{E}_n$ or the asymptotic variance of its estimator is slightly overestimated (or both are). Naturally, there is no evidence whatsoever of an effective coverage smaller than 95% for $\mathcal{E}_n$. The empirical proportions of coverage for $\mathcal{C}_n$, all larger than 98.9% and often equal to 100%, illustrate the fact that the lower confidence bounds are conservative by construction.

Finally, the mean values of $|\mathcal{E}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{E}_n|$ and $|\mathcal{C}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{C}_n|$ quickly stabilize around 1.30. They quantify how close the lower confidence bounds are to the parameters they lower bound, at the scale of the parameters themselves (which, by nature, are bound to get close to zero, if not to converge to it).

6.3 Illustration

Figures 1 and 2 illustrate the data-adaptive inference of the optimal treatment rule, its mean reward and the related pseudo-regrets with a visual summary of one additional run of the procedure described in Sections 2.2 and 5. We see in the top plot of Figure 1 that each 95%-confidence interval contains both its corresponding data-adaptive parameter $\psi_{r_n,0}$ and ψ_0. Moreover, the difference between the length of the 95%-confidence interval at sample size n and that of the vertical segment joining the two grey curves at this sample size gets smaller as n grows, showing that the variance of ψ_n^* gets closer to the optimal variance $\mathrm{Var}_{P_{Q_0,r_0}} D(Q_0, r_0)(O)$. Finally, the bottom plot also reveals that the empirical and counterfactual cumulated pseudo-regrets $\mathcal{C}_n$ and $\mathcal{E}_n$ go to zero and that each 95%-lower confidence-bound is indeed below its corresponding pseudo-regrets.

7 Discussion

We develop a targeted, data-adaptive sampling scheme and TMLE estimator to build confidence intervals on the mean reward under the current estimate of the optimal treatment rule and the optimal treatment rule itself. As a by product, we also obtain lower confidence bounds on two cumulated pseudo-regrets. A simulation study illustrates the theoretical results. One of the cornerstones of the study is a new maximal inequality for martingales wrt the uniform entropy integral which allows the control of several empirical processes indexed by random functions.

We assume here that there is no stratum of the baseline covariates where treatment is neither beneficial nor harmful, *i.e.*, that non-exceptionality holds [21]. In future work, we will extend our result to handle exceptionality, building upon [16] (where observations are sampled independently). Extension to more than two treatments and to the inference of an optimal dynamic treatment rule (where treatment assignment consists in successive assignments at successive time points) and its mean reward will also be considered.

Acknowledgements. Antoine Chambaz and Wenjing Zheng acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-13-BS01-0005 (project SPADRO). Mark J. van der Laan was funded by NIH Grant Number 2R01 A1074345-07.

A Proofs

The notation $a \lesssim b$ means that expression a is smaller than expression b up to a universal multiplicative constant.

To alleviate notation, we introduce the indexing parameter $\zeta \in \cup_{n\geq 1} B_n \times \mathcal{G}_1$ which stands for a couple (β, g). For every $\zeta \equiv (\beta, g) \in \cup_{n\geq 1} B_n \times \mathcal{G}_1$, $\rho \in \mathcal{R}$ and $\epsilon \in \mathcal{E}$, we set

$$Q_{Y,\zeta,\rho}(\epsilon) \equiv \operatorname{expit}\left(\operatorname{logit}(Q_{Y,\beta}) + \epsilon H_\rho(g)\right) \tag{36}$$

and characterize $Q_{Y,\zeta,\rho}(\epsilon) \circ \rho$ given by

$$Q_{Y,\zeta,\rho}(\epsilon) \circ \rho(W) = Q_{Y,\zeta,\rho}(\epsilon)(\rho(W), W).$$

With $\zeta_n \equiv (\beta_n, g_n)$ and $\zeta_0 \equiv (\beta_0, g_0)$, we set

$$\begin{aligned} Q^*_{Y,\zeta_n,r_n} &\equiv Q_{Y,\zeta_n,r_n}(\epsilon_n), \\ Q^*_{Y,\zeta_0,r_n} &\equiv Q_{Y,\zeta_0,r_n}(\epsilon_0(r_n)) \end{aligned}$$

where $\epsilon_0(r_n)$ is defined in (20) with $\rho \equiv r_n$. With both $\zeta = \zeta_n$ and $\zeta = \zeta_0$, we also introduce $Q^*_{Y,\zeta,r_n} \circ r_n$ and d^*_{Y,ζ,r_n} given by

$$Q^*_{Y,\zeta,r_n} \circ r_n(W) \equiv Q^*_{Y,\zeta,r_n}(r_n(W), W), \tag{37}$$

$$d^*_{Y,\zeta,r_n}(O, Z) \equiv \frac{\mathbf{1}\{A = r_n(W)\}}{Z}\left(Y - Q^*_{Y,\zeta,r_n}(A, W)\right). \tag{38}$$

In particular, $d^*_{Y,\zeta_n,r_n} = d^*_{Y,n}$ previously defined in (29). Finally, we denote Q^*_{ζ,r_n} any $Q \in \mathcal{Q}$ such that the marginal distribution of W under Q is the empirical measure and $Q_Y = Q^*_{Y,\zeta,r_n}$.

Lemmas 2, 3 and 4 are proven in Section A.1. Proposition 1 and Theorem 2 in Section A.2 and Propositions 4, 5 and 6 in Section A.3. Technical lemmas are presented and proven in Section B.

A.1 Proofs of Lemmas 2, 3 and 4

Proof of Lemma 2. The key to the proof is the following identity: for each $g \in \mathcal{G}$, we have

$$E_{Q_0,g}(Q_{Y,0}(A, W)) = E_{Q_0}(Q_{Y,0}(0, W)) + E_{Q_0}(q_{Y,0}(W)g(1|W)). \tag{39}$$

This is a straightforward consequence of the decomposition $Q_{Y,0}(A, W) = Q_{Y,0}(0, W) + Aq_{Y,0}(W)$. Moreover, (39) also holds when g takes its value in $[0, 1]$, hence for all treatment rules as well.

Set $n \geq 1$. Applying (39) with $g = r_n$ and $g = r_0$ yields

$$E_{Q_0,r_n}(Q_{Y,0}(A, W)) = E_{Q_0}(Q_{Y,0}(0, W)) + E_{Q_0}(q_{Y,0}(W)r_n(W)), \tag{40}$$

$$E_{Q_0,r_0}(Q_{Y,0}(A, W)) = E_{Q_0}(Q_{Y,0}(0, W)) + E_{Q_0}(q_{Y,0}(W)r_0(W)). \tag{41}$$

Because $E_{Q_0,r_0}(Q_{Y,0}(A, W)) = E_{Q_0}(Q_{Y,0}(r_0(W), W)) = \psi_0$, substracting (40) and (41) entails

$$\psi_0 - E_{Q_0,r_n}(Q_{Y,0}(A, W)) = E_{Q_0}\left(q_{Y,0}(W) \times (r_0(W) - r_n(W))\right) \leq \|r_n - r_0\|_1. \tag{42}$$

By definition of r_0, the above LHS expression is non-negative, hence it coincides with $\Delta(r_n, r_0)$. This completes the proof of (16).

We now apply (39) with $g = g_0$ to get

$$E_{Q_0,g_0}(Q_{Y,0}(A,W)) = E_{Q_0}(Q_{Y,0}(0,W)) + E_{Q_0}(q_{Y,0}(W)g_0(1|W)). \tag{43}$$

Substracting (43) and (41) yields the new equality

$$0 \le \psi_0 - E_{Q_0,g_0}(Q_{Y,0}(A,W)) = E_{Q_0}\left(q_{Y,0}(W) \times (r_0(W) - g_0(1|W))\right).$$

Based on (12), a case-by-case study depending on the sign of $q_{Y,0}(W)$ finally reveals that

$$0 \le \psi_0 - E_{Q_0,g_0}(Q_{Y,0}(A,W)) \le t_\infty E_{Q_0}\left(|q_{Y,0}(W)|\right) + \xi_\infty \le t_\infty + \xi_\infty. \tag{44}$$

To obtain (17), we simply note that

$$\begin{aligned} 0 &\le \psi_0 - E_{Q_0,g_n}(Q_{Y,0}(A,W)) \\ &= \psi_0 - E_{Q_0,g_0}(Q_{Y,0}(A,W)) + E_{Q_0,g_0}(Q_{Y,0}(A,W)) - E_{Q_0,g_n}(Q_{Y,0}(A,W)) \\ &\le t_\infty + \xi_\infty + \Delta(g_n, g_0) \end{aligned}$$

by (44) and (15). □

Proof of Lemma 3. Set $n \ge 1$, $p \ge 1$ and $\eta > 0$. There exists $\alpha > 0$ such that $P_{Q_0}(0 < |q_{Y,0}(W)| < \alpha) \le \eta^p/2$.

Note that $|(r_n - r_0)(W)| \in \{0,1\}$. Moreover, $|(r_n - r_0)(W)| = 1$ implies $q_{Y,\beta_n} q_{Y,0}(W) \le 0$. This justifies the first inequality below. The others easily follow from the fact that $|q_{Y,0}(W)| \le 1$ and a case-by-case study depending on whether $0 < |q_{Y,0}(W)| < \alpha$ or not:

$$\begin{aligned} |q_{Y,0}(W)| \times |(r_n - r_0)(W)|^p &\le |q_{Y,0}(W)| \times \mathbf{1}\{q_{Y,\beta_n} q_{Y,0}(W) \le 0\} \\ &\le \mathbf{1}\{0 < |q_{Y,0}(W)| < \alpha\} + \mathbf{1}\{|q_{Y,0}(W)| \ge \alpha\} \\ &\qquad \times |q_{Y,0}(W)| \times \mathbf{1}\{|(q_{Y,\beta_n} - q_{Y,0})(W)| \ge \alpha\} \\ &\le \mathbf{1}\{0 < |q_{Y,0}(W)| < \alpha\} + \mathbf{1}\{|q_{Y,0}(W)| \ge \alpha\} \\ &\qquad \times |q_{Y,0}(W)| \times \alpha^{-1} |(q_{Y,\beta_n} - q_{Y,0})(W)| \\ &\le \mathbf{1}\{0 < |q_{Y,0}(W)| < \alpha\} \\ &\qquad + \alpha^{-1} |q_{Y,0}(W)|^{1/2} \times |(q_{Y,\beta_n} - q_{Y,0})(W)|. \end{aligned}$$

Taking the expectation under $Q_{W,0} d\mu_W$ on both sides yields

$$\|r_n - r_0\|_p^p \le P_{Q_0}(0 < |q_{Y,0}(W)| < \alpha) + \alpha^{-1} \int |q_{Y,0}|^{1/2} \times |(q_{Y,\beta_n} - q_{Y,0})| Q_{W,0} d\mu_W$$

hence, by choice of α and the Cauchy-Schwartz inequality,

$$\|r_n - r_0\|_p^p \le \eta^p/2 + \alpha^{-1} \|q_{Y,\beta_n} - q_{Y,0}\|_2.$$

Therefore, $\|r_n - r_0\|_p \ge \eta$ implies $\|q_{Y,\beta_n} - q_{Y,0}\|_2 \ge \alpha\eta^p/2$. Consequently, $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1)$ does yield $\|r_n - r_0\|_p = o_P(1)$. This completes the proof. □

Proof of Lemma 4. Set $n \ge 1$, $p \ge 1$, $\bar{p} = p/(p-1)$ ($\bar{p} = \infty$ if $p = 1$) and $p' = \min(p, 2)$.

By (39) with $g = g_n$ and $g = g_0$, we obtain

$$\Delta(g_n, g_0) = |E_{Q_0}[q_{Y,0}(W) \times (g_n(1|W) - g_0(1|W))]|.$$

Applying successively the triangle inequality and Hölder's inequality yields

$$\begin{aligned}\Delta(g_n, g_0) &\leq E_{Q_0}\left(|q_{Y,0}(W)| \times |g_n(1|W) - g_0(1|W)|\right) \\ &\leq \|g_n - g_0\|_p,\end{aligned}$$

which is the result first stated in the lemma.

Suppose now that n is large enough so that $G_n = G_\infty$. Since G_∞ is c_∞-Lipschitz, it holds that

$$\begin{aligned}|q_{Y,0}(W)| \times |g_n(1|W) - g_0(1|W)|^p &= |q_{Y,0}(W)| \times |G_\infty(q_{Y,\beta_n}(W)) - G_\infty(q_{Y,0}(W))|^p \\ &\lesssim |q_{Y,0}(W)| \times |q_{Y,\beta_n}(W) - q_{Y,0}(W)|^p \\ &\leq |q_{Y,0}(W)| \times |q_{Y,\beta_n}(W) - q_{Y,0}(W)|^{p'},\end{aligned}$$

where the last inequality is due to the fact that $|q_{Y,\beta_n} - q_{Y,0}| \leq 1$. Taking the expectation under $Q_{W,0} d\mu_W$ gives the bound $\|g_n - g_0\|_p \lesssim \|q_{Y,\beta_n} - q_{Y,0}\|_{p'}^{p'/p} \lesssim \|q_{Y,\beta_n} - q_{Y,0}\|_2^{p'/p}$. This completes the proof. □

A.2 Proofs of Proposition 1 and Theorem 2

Let us prove Proposition 1.

Proof of Proposition 1. The convergence $\|q_{Y,\beta_n} - q_{Y,\beta_0}\| = o_P(1)$ follows immediately from (18) and the convergence $\|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$. This convergence is a consequence of Lemma 7 with $\Theta \equiv \mathcal{Q}_1$, $\Theta_n \equiv \mathcal{Q}_{1,n}$, d the distance induced on Θ by the norm $\|\cdot\|_{2,P_{Q_0,g^{\text{ref}}}}$, $\mathcal{M}_n$ and $\mathbf{M}_n$ characterized over Θ by $\mathcal{M}_n(Q_Y) \equiv P_{Q_0,g^{\text{ref}}} L(Q_Y)$ (which does not depend on n after all) and $\mathbf{M}_n(Q_Y) \equiv P_n g^{\text{ref}} L(Q_Y)/Z = n^{-1} \sum_{i=1}^n g^{\text{ref}}(A_i \mid W_i) L(Q_Y)(O_i)/Z_i$. Assumption **A2** implies that ***(a)*** and ***(b)*** from Lemma 7 are met (take $\tau_n = Q_{Y,\beta_0}$ and $\tau_n^* = Q_{Y,\beta_{n,0}}$). It remains to prove that ***(c)*** also holds or, in other terms, that $\|\mathbf{M}_n - \mathcal{M}_n\|_{\mathcal{Q}_{1,n}} = o_P(1)$.

For any $Q_Y \in \Theta$, characterize $\ell(Q_Y)$ by setting $\ell(Q_Y)(O, Z) \equiv g^{\text{ref}}(A|W) L(Q_Y)(O)/Z$. Then we can rewrite $\|\mathbf{M}_n - \mathcal{M}_n\|_{\mathcal{Q}_{1,n}}$ as follows:

$$\|\mathbf{M}_n - \mathcal{M}_n\|_{\mathcal{Q}_{1,n}} = \|P_n \ell - P_{Q_0,g^{\text{ref}}} L\|_{\mathcal{Q}_{1,n}} = \|(P_n - P_{Q_0,\mathbf{g}_n})\ell\|_{\mathcal{Q}_{1,n}} = \|P_n - P_{Q_0,\mathbf{g}_n}\|_{\ell(\mathcal{Q}_{1,n})}.$$

The separability of $\ell(\mathcal{Q}_{1,n})$ follows from that of $L(\mathcal{Q}_{1,n})$. Let F_n be the envelope function for $L(\mathcal{Q}_{1,n})$ from **A4**. By construction of g_n, Z is bounded away from 0, so there exists a constant $c > 0$ such that cF_n is an envelope function for $\ell(\mathcal{Q}_{1,n})$. Moreover, $J_{cF_n}(1, \ell(\mathcal{Q}_{1,n})) = O(J_{F_n}(1, L(\mathcal{Q}_{1,n})) = o(\sqrt{n})$ by **A4**. Therefore, Lemma 9 applies and yields $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\ell(\mathcal{Q}_{1,n})} = o_P(1)$ by Markov's inequality. Thus, we can apply Lemma 7. It yields that $\|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$, which is the desired result.

Assume now that **A1** and **A5** also hold and set arbitrarily $t > 0$. Because $|r_n - r_0| \in \{0, 1\}$, we can upper-bound $\|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}}^2$ as follows:

$$\|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}}^2 = P_{Q_0,g^{\text{ref}}} \mathbf{1}\{|q_{Y,0}| > t\} \times |r_n - r_0| + P_{Q_0,g^{\text{ref}}} \mathbf{1}\{|q_{Y,0}| \leq t\} \times |r_n - r_0|$$

$$
\begin{aligned}
&\leq \quad t^{-1} P_{Q_0,g^{\mathrm{ref}}} |q_{Y,0}| \times |r_n - r_0| + P_{Q_0,g^{\mathrm{ref}}}(0 < |q_{Y,0}| \leq t) \\
&\lesssim \quad t^{-1} \|r_n - r_0\|_2^2 + t^{\gamma_2}.
\end{aligned}
$$

Optimizing in t yields

$$
\|r_n - r_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \lesssim \|r_n - r_0\|_2^{\gamma_2/2(1+\gamma_2)} = o_P(1).
$$

We obtain that

$$
\|g_n - g_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \lesssim \|g_n - g_0\|_2^{\gamma_2/2(1+\gamma_2)} = o_P(1)
$$

along the same lines as above. This completes the proof. □

We now turn to the first part of Theorem 2:

Proposition 7 (consistency of ψ_n^*). *Suppose that* **A2**, **A3** *and* **A4** *are met. Then it holds that* $\psi_n^* - \psi_{r_n,0} = o_P(1)$.

Proof of Proposition 7. This is a three-part proof.

Step one: studying ϵ_n. Let us show that $\epsilon_n - \epsilon_0(r_n) = o_P(1)$. We apply Lemma 8 with $\Theta \equiv \mathcal{E}$, d the Euclidean distance, $\mathcal{Z}_n$ and $\mathbf{Z}_n$ characterized over $\mathcal{E}$ by $\mathcal{Z}_n(\epsilon) = P_{Q_0,g_0} D_{Y,r_n}(Q_{Y,\zeta_0,r_n}(\epsilon), g_0)$, and $\mathbf{Z}_n(\epsilon) = P_n D_{Y,r_n}(Q_{Y,\zeta_n,r_n}(\epsilon), g_n) g_n / Z$, see (36), (19) and (8) for the definitions of $Q_{Y,\zeta_0,r_n}(\epsilon)$ and $Q_{Y,\zeta_n,r_n}(\epsilon)$.

From the differentiability of $\epsilon \mapsto L^{\mathrm{kl}}(Q_{Y,\zeta_0,r_n}(\epsilon))$, validity of the differentiation under the integral sign, and definition of $\epsilon_0(r_n)$ (20) in **A3**, we deduce that $\mathcal{Z}_n(\epsilon_0(r_n)) = 0$. By definition of ϵ_n (9), $\mathbf{Z}_n(\epsilon_n) = 0$ too. Moreover, ***(d)*** from Lemma 8 is met. Indeed, by differentiability of $\epsilon \mapsto D_{Y,r_n}(Q_{Y,\zeta_0,r_n}(\epsilon), g_0)$ and validity of the differentiation under the integral sign, $\mathcal{Z}_n : \mathcal{E} \to \mathbb{R}$ is differentiable on $\mathcal{E}$ with a derivative given by

$$
\mathcal{Z}_n'(\varepsilon) = -P_{Q_0,g_0} \frac{Q_{Y,\zeta_0,r_n}(\epsilon) \circ r_n \times (1 - Q_{Y,\zeta_0,r_n}(\epsilon) \circ r_n)}{g_0 \circ r_n}
$$

where $g_0 \circ r_n$ is characterized by $g_0 \circ r_n(W) = g_0(r_n(W)|W)$. By construction, $Q_{Y,\zeta,r}(\epsilon)$ and g_0 are bounded away from 0 and 1 uniformly in $\zeta \in \cup_{n \geq 1} B_n \times \mathcal{G}_1$, $\rho \in \mathcal{R}$ and $\epsilon \in \mathcal{E}$. Therefore, there exists a universal constant c such that $|\mathcal{Z}_n'(\varepsilon)| \geq c > 0$ for all $\epsilon \in \mathcal{E}$. Consequently, by the mean value theorem, for all $\varepsilon \in \mathcal{E}$, $|\mathcal{Z}_n(\epsilon)| \geq c|\epsilon - \epsilon_0(r_n)|$. This entails condition ***(d)***.

Applying Lemma 8 finally requires verifying that ***(e)*** is met, *i.e.*, proving that $\|\mathbf{Z}_n - \mathcal{Z}_n\|_{\mathcal{E}} = o_P(1)$. Introduce $\mathcal{F}_n \equiv \{f_{\rho,\epsilon} : \rho \in r(\mathcal{Q}_{1,n}), \epsilon \in \mathcal{E}\}$ with

$$
f_{\rho,\epsilon}(O, Z) \equiv \frac{\mathbf{1}\{A = \rho(W)\}}{Z} (Y - Q_{Y,\zeta_0,\rho}(\epsilon)(A, W)) \tag{45}
$$

for each $(\rho, \epsilon) \in r(\mathcal{Q}_{1,n}) \times \mathcal{E}$. We start with the following derivation:

$$
\begin{aligned}
\|\mathbf{Z}_n(\epsilon) - &\mathcal{Z}_n(\epsilon)\|_{\mathcal{E}} \\
&= \sup_{\epsilon \in \mathcal{E}} \left| P_n \left(f_{r_n,\epsilon} + \frac{\mathbf{1}\{A = r_n(W)\}}{Z} (Q_{Y,\zeta_0,r_n}(\epsilon) - Q_{Y,\zeta_n,r_n}(\epsilon)) \right) - P_{Q_0,\mathbf{g}_n} f_{r_n,\epsilon} \right| \\
&\leq \|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}_n} + \sup_{\epsilon \in \mathcal{E}} \left| P_n \frac{\mathbf{1}\{A = r_n(W)\}}{Z} (Q_{Y,\zeta_0,r_n}(\epsilon) - Q_{Y,\zeta_n,r_n}(\epsilon)) \right|. \qquad (46)
\end{aligned}
$$

- Consider the first RHS term in (46). Set $(\rho_1, \epsilon_1), (\rho_2, \epsilon_2) \in r(\mathcal{Q}_{1,n}) \times \mathcal{E}$. Since Z is bounded away from 0 and $|Y - Q_{Y,\zeta_0,\rho_2}(\epsilon_2)(A, W)| \leq 1$, it holds that

$$\begin{aligned}
|(f_{\rho_1,\epsilon_1} - f_{\rho_2,\epsilon_2})(O, Z)| &\leq \frac{\mathbf{1}\{A = \rho_1(W)\}}{Z} \left|\left(Q_{Y,\zeta_0,\rho_1}(\epsilon_1) - Q_{Y,\zeta_0,\rho_2}(\epsilon_2)\right)(A, W)\right| \\
&\quad + |\mathbf{1}\{A = \rho_1(W) - \mathbf{1}\{A = \rho_2(W)\}| \\
&\quad\quad \times \frac{|Y - Q_{Y,\zeta_0,\rho_2}(\epsilon_2)(A, W)|}{Z} \\
&\lesssim \left|\left(Q_{Y,\zeta_0,\rho_1}(\epsilon_1) - Q_{Y,\zeta_0,\rho_1}(\epsilon_2)\right)(A, W)\right| \\
&\quad + \left|\left(Q_{Y,\zeta_0,\rho_1}(\epsilon_2) - Q_{Y,\zeta_0,\rho_2}(\epsilon_2)\right)(A, W)\right| \\
&\quad + |\rho_1(W) - \rho_2(W)|.
\end{aligned}$$

  Because expit is 1-Lipschitz, $\mathcal{E}$ is bounded and g_0 is bounded away from 0, this entails the bound

$$\begin{aligned}
|(f_{\rho_1,\epsilon_1} - f_{\rho_2,\epsilon_2})(O, Z)| &\lesssim |\epsilon_1 - \epsilon_2| + |\epsilon_2| \times |(H_{\rho_1}(g_0) - H_{\rho_2}(g_0))(O)| \\
&\quad + |\rho_1(W) - \rho_2(W)| \\
&\lesssim |\epsilon_1 - \epsilon_2| + |\rho_1(W) - \rho_2(W)|. \qquad (47)
\end{aligned}$$

  This upper-bound notably implies that $\mathcal{F}_n$ is separable because $r(\mathcal{Q}_{1,n})$ and $\mathcal{E}$ (seen as a class of constant functions) are separable. By **A4**, $J_1(1, r(\mathcal{Q}_{1,n})) = o(\sqrt{n})$. Since $\mathcal{E}$ is bounded, there exists a bounded envelope function F for $\mathcal{E}$ seen as a class of (constant) functions and $J_F(1, \mathcal{E})$ is finite. Assume without loss of generality that F is also an envelope function for $\mathcal{F}_n$. By (47) and the trivial inequalities $(a+b)^2 \leq 2(a^2+b^2)$ and $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$ (valid for all $a, b \geq 0$), $J_F(1, \mathcal{F}_n) = o(\sqrt{n})$ (we will use repeatedly this argument in the rest of the article, without mentioning its details). Therefore, we can apply Lemma 9 and conclude, with Markov's inequality, that $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}_n} = o_P(1)$.

- Consider next the second term in the RHS of (46). It is upper-bounded by

$$\Delta_n \equiv \sup_{\epsilon \in \mathcal{E}} P_n |Q_{Y,\zeta_0,r_n}(\epsilon) - Q_{Y,\zeta_n,r_n}(\epsilon)|/Z.$$

  Since expit is 1-Lipschitz, $\mathcal{Q}_{1,n}$ is bounded away from 0 and 1, and logit is Lipschitz on any compact subset of $]0, 1[$, it holds that

$$\begin{aligned}
\Delta_n &\leq \sup_{\epsilon \in \mathcal{E}} P_n \left|\text{logit}(Q_{Y,\beta_n}) - \text{logit}(Q_{Y,\beta_0}) + \epsilon(H_{r_n}(g_n) - H_{r_n}(g_0))\right|/Z \\
&\lesssim P_n|Q_{Y,\beta_n} - Q_{Y,\beta_0}|/Z + P_n|1/g_n - 1/g_0|/Z \\
&= P_{Q_0,\mathbf{g}_n}|Q_{Y,\beta_n} - Q_{Y,\beta_0}|/Z + P_{Q_0,\mathbf{g}_n}|1/g_n - 1/g_0|/Z \\
&\quad + (P_n - P_{Q_0,\mathbf{g}_n})|Q_{Y,\beta_n} - Q_{Y,\beta_0}|/Z + (P_n - P_{Q_0,\mathbf{g}_n})|1/g_n - 1/g_0|/Z. \qquad (48)
\end{aligned}$$

  Using the fact that g^{ref} is bounded away from 0 and 1 and the Cauchy-Schwarz inequality, we readily see that $P_{Q_0,\mathbf{g}_n}|Q_{Y,\beta_n} - Q_{Y,\beta_0}|/Z \lesssim P_{Q_0,g^{\text{ref}}}|Q_{Y,\beta_n} - Q_{Y,\beta_0}| \leq \|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ by Proposition 1, whose assumptions are met here. We control $P_{Q_0,\mathbf{g}_n}|1/g_n - 1/g_0|/Z$ similarly, using additionally that $\mathbf{g}_n$ and g_0 are uniformly bounded away from 0 and 1 and that, for n large enough, $G_n = G_\infty$ is c_∞-Lipschitz. Indeed, for n large enough, $P_{Q_0,\mathbf{g}_n}|1/g_n - 1/g_0|/Z \lesssim P_{Q_0,g^{\text{ref}}}|g_n - g_0| \leq \|g_n - g_0\|_{2,P_{Q_0,g^{\text{ref}}}}$ and

$$\|g_n - g_0\|_{2,P_{Q_0,g^{\text{ref}}}} = \|G_\infty(q_{Y,\beta_n}) - G_\infty(q_{Y,\beta_0})\|_{2,P_{Q_0,g^{\text{ref}}}}$$

$$\begin{aligned} &\lesssim \; \|q_{Y,\beta_n} - q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \\ &\lesssim \; \|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1), \end{aligned} \tag{49}$$

as recalled earlier. Thus, the sum of the two first terms in the RHS expression of (48) is $o_P(1)$.

We now turn to the third term of the RHS sum in (48). For any $Q_Y \in \mathcal{Q}_1$, introduce $h_1(Q_Y)$ characterized by $h_1(Q_Y)(O,Z) \equiv |Q_Y(A,W) - Q_{Y,\beta_0}(A,W)|/Z$. Obviously,

$$|(P_n - P_{Q_0,\mathbf{g}_n})|Q_{Y,\beta_n} - Q_{Y,\beta_0}|/Z| \leq \|(P_n - P_{Q_0,\mathbf{g}_n})h_1\|_{\mathcal{Q}_{1,n}} = \|P_n - P_{Q_0,\mathbf{g}_n}\|_{h_1(\mathcal{Q}_{1,n})}.$$

The separability of $\mathcal{Q}_{1,n}$ implies that of $h_1(\mathcal{Q}_{1,n})$. Since Z is bounded away from 0, it holds that $h_1(\mathcal{Q}_1)$ is uniformly bounded by a constant $c > 0$ which can serve as a constant envelope function, and $J_c(1, h_1(\mathcal{Q}_{1,n})) = O(J_1(1, \{|Q_Y - Q_{Y,\beta_0} : Q_Y \in \mathcal{Q}_{1,n}|\})) = O(J_1(1, \mathcal{Q}_{1,n})) = o(\sqrt{n})$ by **A4**. Therefore, Lemma 9 applies and Markov's inequality yields $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{h_1(\mathcal{Q}_{1,n})} = o_P(1)$. We control the last term similarly. Let n be large enough so that $G_n = G_\infty$. For any $Q_Y \in \mathcal{Q}_1$, introduce $h_2(Q_Y)$ characterized by $h_2(Q_Y)(O,Z) \equiv |1/G_\infty(q_Y(A,W)) - 1/G_\infty(q_{Y,\beta_0}(A,W))|/Z$. We have

$$|(P_n - P_{Q_0,\mathbf{g}_n})|1/g_n - 1/g_0|/Z| \leq \|(P_n - P_{Q_0,\mathbf{g}_n})h_2\|_{\mathcal{Q}_{1,n}} = \|P_n - P_{Q_0,\mathbf{g}_n}\|_{h_2(\mathcal{Q}_{1,n})}.$$

The separability of $\mathcal{Q}_{1,n}$ implies that of $h_2(\mathcal{Q}_{1,n})$. Because Z is bounded away from 0 and because G_∞ is c_∞-Lipschitz and bounded away from 0 and 1 too, it holds that $h_2(\mathcal{Q}_1)$ is uniformly bounded by a constant $c' > 0$ which can serve as a constant envelope function, and $J_{c'}(1, h_2(\mathcal{Q}_{1,n})) = O(J_1(1, \{|q_Y - q_{Y,\beta_0}| : Q_Y \in \mathcal{Q}_{1,n}\})) = O(J_1(1, \{|Q_Y - Q_{Y,\beta_0}| : Q_Y \in \mathcal{Q}_{1,n}\})) = O(J_1(1, \mathcal{Q}_{1,n})) = o(\sqrt{n})$, as we have seen before. Thus, $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{h_2(\mathcal{Q}_{1,n})} = o_P(1)$, hence the sum of the tow last terms in the RHS expression of (48) is $o_P(1)$. We conclude that $\Delta_n = o_P(1)$.

Combining the results obtained on the first and second RHS terms in (46) yields the desired convergence $\|\mathbf{Z}_n - \mathcal{Z}_n\|_{\mathcal{E}} = o_P(1)$. We are now in a position to apply Lemma 7, which implies the stated convergence $\epsilon_n - \epsilon_0(r_n) = o_P(1)$.

Step two: studying Q^*_{Y,ζ_n,r_n}. Let us now prove that $\|Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1)$. For this, we equip $\mathcal{Q}_1 \times \mathcal{G}_1 \times \mathcal{E} - \mathcal{Q}_1 \times \mathcal{G}_1 \times \mathcal{E}$ with a seminorm $|||\cdot|||_1$ such that, for any two $(Q_{Y,1}, g_1, \epsilon_1), (Q_{Y,2}, g_2, \epsilon_2) \in \mathcal{Q}_1 \times \mathcal{G}_1 \times \mathcal{E}$,

$$|||(Q_{Y,1}, g_1, \epsilon_1) - (Q_{Y,2}, g_2, \epsilon_2)|||_1 \equiv \|Q_{Y,1} - Q_{Y,2}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + \|g_1 - g_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2|.$$

Proposition 1 and the first step of this proof imply that

$$|||(Q_{Y,\beta_n}, g_n, \epsilon_n) - (Q_{Y,\beta_0}, g_0, \epsilon_0(r_n))|||_1 = o_P(1).$$

We also equip the set $\mathcal{Q}_Y^{\mathcal{R}} - \mathcal{Q}_Y^{\mathcal{R}}$ with a seminorm $|||\cdot|||_2$ characterized as follows: for any two $(Q_{Y,\rho})_{\rho\in\mathcal{R}}, (Q'_{Y,\rho})_{\rho\in\mathcal{R}} \in \mathcal{Q}_Y^{\mathcal{R}}$,

$$\left|\left|\left|(Q_{Y,\rho})_{\rho\in\mathcal{R}} - (Q'_{Y,\rho})_{\rho\in\mathcal{R}}\right|\right|\right|_2 \equiv \sup_{\rho\in\mathcal{R}} \|Q_{Y,\rho} - Q'_{Y,\rho}\|_{2,P_{Q_0,g^{\mathrm{ref}}}}.$$

Let $\boldsymbol{f} : \mathcal{Q}_1 \times \mathcal{G}_1 \times \mathcal{E} \to \mathcal{Q}_Y^{\mathcal{R}}$ be given by $\boldsymbol{f}(Q_Y, g, \epsilon) = (f_\rho(Q_Y, g, \epsilon))_{\rho\in\mathcal{R}}$ where, for each $\rho \in \mathcal{R}$,

$$f_\rho(Q_Y, g, \epsilon)(O) \equiv \mathrm{expit}\left(\mathrm{logit}(Q_Y(A,W)) + \epsilon H_\rho(g)(O)\right). \tag{50}$$

Set $(Q_{Y,1}, g_1, \epsilon_1), (Q_{Y,2}, g_2, \epsilon_2) \in \mathcal{Q}_1 \times \mathcal{G}_1 \times \mathcal{E}$ and $\rho \in \mathcal{R}$. Because *(i)* expit is 1-Lipschitz, *(ii)* $\mathcal{Q}_1$ is bounded away from 0 and 1, and logit is Lipschitz on any compact subset of $]0, 1[$, *(iii)* $\mathcal{G}_1$ is uniformly bounded away from 0 and 1, *(iv)* $\mathcal{E}$ is a bounded set, it holds that

$$\begin{aligned}
&\|f_\rho(Q_{Y,1}, g_1, \epsilon_1) - f_\rho(Q_{Y,2}, g_2, \epsilon_2)\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \\
&\quad \leq \|\operatorname{logit}(Q_{Y,1}) - \operatorname{logit}(Q_{Y,2})\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + \|\epsilon_2(1/g_1 - 1/g_2)\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \\
&\qquad + \|(\epsilon_1 - \epsilon_2)/g_1\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \\
&\quad \lesssim \|Q_{Y,1} - Q_{Y,2}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + \|g_1 - g_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2| \\
&\quad = \|\!|(Q_{Y,1}, g_1, \epsilon_1) - (Q_{Y,2}, g_2, \epsilon_2)|\!\|_1. \qquad (51)
\end{aligned}$$

Noting that the RHS expression does not depend on ρ then taking the supremum in $\rho \in \mathcal{R}$ to the left yields

$$\|\!|\boldsymbol{f}(Q_{Y,1}, g_1, \epsilon_1) - \boldsymbol{f}(Q_{Y,2}, g_2, \epsilon_2)|\!\|_2 \lesssim \|\!|(Q_{Y,1}, g_1, \epsilon_1) - (Q_{Y,2}, g_2, \epsilon_2)|\!\|_1.$$

Therefore, the convergence $\|\!|(Q_{Y,\beta_n}, g_n, \epsilon_n) - (Q_{Y,\beta_0}, g_0, \epsilon_0(r_n))|\!\|_1 = o_P(1)$ implies the convergence $\|\!|\boldsymbol{f}(Q_{Y,\beta_n}, g_n, \epsilon_n) - \boldsymbol{f}(Q_{Y,\beta_0}, g_0, \epsilon_0(r_n))|\!\|_2 = o_P(1)$. In particular,

$$\|f_{r_n}(Q_{Y,\beta_n}, g_n, \epsilon_n) - f_{r_n}(Q_{Y,\beta_0}, g_0, \epsilon_0(r_n))\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = \|Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1),$$

as we claimed.

Step three: studying ψ_n^*. Let us first demonstrate that $E_{Q_{W,0}}(Q^*_{Y,\zeta_0,r_n} \circ r_n(W)) = \psi_{r_n,0}$, then that $\psi_n^* - \psi_{r_n,0} = o_P(1)$. We have already shown that $\mathcal{Z}_n(\epsilon_0(r_n)) = 0$. Equivalently, by conditioning first on (A, W) (second line) then on W only (third line),

$$\begin{aligned}
0 = \mathcal{Z}_n(\epsilon_0(r_n)) &= E_{Q_0,g_0}\left(\frac{\mathbf{1}\{A = r_n(W)\}}{g_0(A|W)}(Y - Q^*_{Y,\zeta_0,r_n}(A, W))\right) \\
&= E_{Q_0,g_0}\left(\frac{\mathbf{1}\{A = r_n(W)\}}{g_0(A|W)}(Q_{Y,0}(A, W) - Q^*_{Y,\zeta_0,r_n}(A, W))\right) \\
&= E_{Q_0,g_0}\left(Q_{Y,0}(r_n(W), W) - Q^*_{Y,\zeta_0,r_n}(r_n(W), W)\right) \\
&= \psi_{r_n,0} - E_{Q_{W,0}}(Q^*_{Y,\zeta_0,r_n} \circ r_n(W)) \qquad (52)
\end{aligned}$$

hence the claimed equality.

Let $\psi_n^\sim \equiv E_{Q_{W,0}}(Q^*_{Y,\zeta_n,r_n} \circ r_n(W))$. By (52), the fact that g^{ref} is bounded away from 0 and 1 and the Cauchy-Schwarz inequality, it holds that

$$\begin{aligned}
|\psi_n^\sim - \psi_{r_n,0}| &= \left|E_{Q_{W,0},g^{\mathrm{ref}}}\left(\frac{\mathbf{1}\{A = r_n(W)\}}{g^{\mathrm{ref}}(A|W)}\left(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}\right)(A, W)\right)\right| \\
&\lesssim \|Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1). \qquad (53)
\end{aligned}$$

Therefore, it suffices to show that $\psi_n^* - \psi_n^\sim = o_P(1)$ too to conclude.

Since $Q^*_{Y,\zeta_n,r_n} \circ r_n$ is a function of W only, we can write

$$|\psi_n^* - \psi_n^\sim| = \left|(P_n - P_{Q_0,\mathbf{g}_n})Q^*_{Y,\zeta_n,r_n} \circ r_n\right| \leq \|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}'_n}$$

where we define $\mathcal{F}'_n \equiv \{Q_{Y,\zeta,\rho}(\epsilon) \circ \rho : \zeta \in B_n \times \mathcal{G}_{1,n}, \rho \in r(\mathcal{Q}_{1,n}), \epsilon \in \mathcal{E}\}$. By construction, $\mathcal{F}'_n$ is uniformly bounded by 1 which can serve as an envelope function. Moreover, for every $\zeta_1 \equiv (\beta_1, g_1), \zeta_2 \equiv (\beta_2, g_2) \in B_n \times \mathcal{G}_{1,n}, \rho_1, \rho_2 \in r(\mathcal{Q}_{1,n}), \epsilon_1, \epsilon_2 \in \mathcal{E}$, because *(i)* $|(\rho_1 - \rho_2)(W)| \in \{0, 1\}$, *(ii)* expit is 1-Lipschitz, *(iii)* $\mathcal{Q}_1$ is bounded away from 0 and 1, logit is Lipschitz on any compact subset of $]0,1[$, and *(iv)* $\mathcal{G}_1$ is uniformly bounded away from 0 and 1, the following inequalities hold pointwise:

$$
\begin{aligned}
&|Q_{Y,\zeta_1,\rho_1}(\epsilon_1) \circ \rho_1 - Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_2| \\
&\quad = \left|\left(Q_{Y,\zeta_1,\rho_1}(\epsilon_1) - Q_{Y,\zeta_2,\rho_2}(\epsilon_2)\right) \circ \rho_1 + \left(Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_1 - Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_2\right)\right| \\
&\quad \leq \left|\left(Q_{Y,\zeta_1,\rho_1}(\epsilon_1) - Q_{Y,\zeta_2,\rho_2}(\epsilon_2)\right) \circ \rho_1\right| \\
&\qquad + |\rho_1 - \rho_2| \left|\left(Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_1 - Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_2\right)\right| \\
&\quad \lesssim |(Q_{Y,\beta_1} - Q_{Y,\beta_2}) \circ \rho_1| + |\epsilon_1/g_1(\rho_1|\cdot) - \epsilon_2/g_2(\rho_1|\cdot)| + |\rho_1 - \rho_2| \\
&\quad \lesssim |(Q_{Y,\beta_1} - Q_{Y,\beta_2}) \circ \rho_1| + |g_1(\rho_1|\cdot) - g_2(\rho_1|\cdot)| + |\epsilon_1 - \epsilon_2| + |\rho_1 - \rho_2| \\
&\quad \leq |(Q_{Y,\beta_1} - Q_{Y,\beta_2}) \circ \rho_1| + |(Q_{Y,\beta_1} - Q_{Y,\beta_2}) \circ (1 - \rho_1)| + |g_1(\rho_1|\cdot) - g_2(\rho_1|\cdot)| \\
&\qquad + |g_1(1 - \rho_1|\cdot) - g_2(1 - \rho_1|\cdot)| + |\epsilon_1 - \epsilon_2| + |\rho_1 - \rho_2| \\
&\quad = |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| + 2|g_1 - g_2| + |\epsilon_1 - \epsilon_2| + |\rho_1 - \rho_2| \qquad (54)
\end{aligned}
$$

where, for every $\beta \in \cup_{n\geq 1} B_n$, $Q^-_{Y,\beta}$ denotes the function given by $Q^-_{Y,\beta}(A, W) \equiv Q_{Y,\beta}(1 - A, W)$. This entails that $\mathcal{F}'_n$ is separable because $\mathcal{Q}_{1,n}$, $\mathcal{G}_{1,n}$, $\mathcal{E}$ (seen as a class of constant functions with envelope function $F' \geq 1$) and $r(\mathcal{Q}_{1,n})$ are separable (the separability of $\mathcal{G}_{1,n}$ follows straightforwardly from that of $\mathcal{Q}_{1,n}$, the definition of $\mathcal{G}_{1,n}$ and continuity of G_n). Let n be large enough so that $G_n = G_\infty$. Inequality (54) and the facts that *(i)* $\mathcal{G}_{1,n} \equiv \{G_n(q_Y) : Q_Y \in \mathcal{Q}_{1,n}\} = \{G_\infty(q_Y) : Q_Y \in \mathcal{Q}_{1,n}\}$ with G_∞ c_∞-Lipschitz and *(ii)* $|q_{Y,\beta_1} - q_{Y,\beta_2}| \leq |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}|$ imply that

$$
\begin{aligned}
J_{F'}(1, \mathcal{F}'_n) &\lesssim J_{F'}(1, \mathcal{Q}_{1,n}) + J_{F'}(1, \mathcal{G}_{1,n}) + J_{F'}(1, \mathcal{E}) + J_{F'}(1, r(\mathcal{Q}_{1,n})) \\
&\lesssim J_{F'}(1, \mathcal{Q}_{1,n}) + J_{F'}(1, \mathcal{E}) + J_{F'}(1, r(\mathcal{Q}_{1,n})) = o(\sqrt{n})
\end{aligned}
$$

by **A4**. Thus, Lemma 9 applies and Markov's inequality yields $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}'_n} = o_P(1)$ hence $|\psi^*_n - \psi^\sim_n| = o_P(1)$. This completes the third step, and the proof of Proposition 7. □

The second part of Theorem 2 revolves around a consequence of the following result.

Proposition 8 (first asymptotic linear expansion of ψ^*_n)**.** *Suppose that* **A2***,* **A3***,* **A4** *and* **A4*** *are met. Then it holds that*

$$
\psi^*_n - \psi_{r_n,0} = (P_n - P_{Q_0,\mathbf{g}_n})(d^*_{Y,\zeta_0,r_n} + D_{W,r_n}(Q^*_{\zeta_0,r_n})) + o_P(1/\sqrt{n}). \qquad (55)
$$

The asymptotic linear expansion (55) is obtained from the exact linear expansion stated in the next lemma.

Lemma 5 (exact linear expansion of ψ^*_n)**.** *It follows from the definition of* ψ^*_n *that*

$$
\begin{aligned}
\psi^*_n - \psi_{r_n,0} &= -P_{Q_0,g_0} D_{r_n}(Q^*_{\zeta_n,r_n}, g_0) \qquad (56)\\
&= (P_n - P_{Q_0,\mathbf{g}_n})(d^*_{Y,\zeta_0,r_n} + D_{W,r_n}(Q^*_{\zeta_0,r_n})) \\
&\quad + (P_n - P_{Q_0,\mathbf{g}_n})\left((d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n})\right. \\
&\qquad\qquad \left. + (Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n\right). \qquad (57)
\end{aligned}
$$

Proof of Lemma 5. Equality (56) readily follows from the definitions of $D_{r_n}(Q^*_{\zeta_n,r_n}, g_0)$, ψ^*_n and $\psi_{r_n,0}$.

We now turn to (57). Let us denote by $P_{n,\mathbf{g}_n}$ the empirical distribution of $\mathbf{O}_n$ weighted by $g_n(A_i|W_i)/g_i(A_i|W_i)$. By construction of the fluctuation (8) and definition of ϵ_n (9), it holds that $P_{n,\mathbf{g}_n} D_{Y,r_n}(Q^*_{Y,\zeta_n,r_n}, g_n) = 0$. Moreover, (10) can be rewritten as $P_n D_{W,r_n}(Q^*_{\zeta_n,r_n}) = 0$. Therefore, (56) is equivalent to

$$\begin{aligned}\psi_n^* - \psi_{r_n,0} = (P_n - P_{Q_0,g_0}) D_{W,r_n}(Q^*_{\zeta_n,r_n}) \\ + \left(P_{n,\mathbf{g}_n} D_{Y,r_n}(Q^*_{Y,\zeta_n,r_n}, g_n) - P_{Q_0,g_0} D_{Y,r_n}(Q^*_{Y,\zeta_n,r_n}, g_0)\right). \quad (58)\end{aligned}$$

Adding and substracting $(P_n - P_{Q_0,g_0}) D_{W,r_n}(Q^*_{\zeta_0,r_n})$ to the first term in the RHS expression of (58) implies

$$\begin{aligned}(P_n - &P_{Q_0,g_0}) D_{W,r_n}(Q^*_{\zeta_n,r_n}) \\ &= (P_n - P_{Q_0,g_0}) D_{W,r_n}(Q^*_{\zeta_0,r_n}) + (P_n - P_{Q_0,g_0})(D_{W,r_n}(Q^*_{\zeta_n,r_n}) - D_{W,r_n}(Q^*_{\zeta_0,r_n})) \\ &= (P_n - P_{Q_0,g_0}) D_{W,r_n}(Q^*_{\zeta_0,r_n}) + (P_n - P_{Q_0,g_0})(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n \\ &= (P_n - P_{Q_0,g_n}) D_{W,r_n}(Q^*_{\zeta_0,r_n}) + (P_n - P_{Q_0,g_n})(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n, \quad (59)\end{aligned}$$

where the last equality is valid because $(P_n - P_{Q_0,g_0})$ operates on functions of W.

As for the second term in the RHS expression of (58), it equals

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \Bigg(&\frac{g_n(A_i|W_i)}{g_i(A_i|W_i)} \frac{\mathbf{1}\{A_i = r_n(W_i)\}}{g_n(A_i|W_i)} (Y_i - Q^*_{Y,\zeta_n,r_n}(A_i, W_i)) \\ &- P_{Q_0,g_0} \frac{\mathbf{1}\{A = r_n(W)\}}{g_0(A|W)} (Y - Q^*_{Y,\zeta_n,r_n})\Bigg) \\ = \frac{1}{n}\sum_{i=1}^n \Bigg(&\frac{\mathbf{1}\{A_i = r_n(W_i)\}}{g_i(A_i \mid W_i)} (Y_i - Q^*_{Y,\zeta_n,r_n}(A_i, W_i)) \\ &- P_{Q_0,g_i} \frac{\mathbf{1}\{A = r_n(W)\}}{g_i(A \mid W)} (Y - Q^*_{Y,\zeta_n,r_n})\Bigg) \\ = (P_n - &P_{Q_0,\mathbf{g}_n}) d^*_{Y,\zeta_n,r_n} \\ = (P_n - &P_{Q_0,\mathbf{g}_n}) d^*_{Y,\zeta_0,r_n} + (P_n - P_{Q_0,\mathbf{g}_n})(d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n}). \quad (60)\end{aligned}$$

The equalities (58), (59) and (60) imply (57). □

We now build upon Lemma 5 to prove Proposition 8.

Proof of Proposition 8. By (57) in Lemma 5, (55) follows from

$$\begin{aligned}(P_n - P_{Q_0,\mathbf{g}_n}) &\left((d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n}) + (Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n\right) \\ &= (P_n - P_{Q_0,\mathbf{g}_n})(d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n}) \\ &\quad + (P_n - P_{Q_0,\mathbf{g}_n})(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n = o_P(1/\sqrt{n}). \quad (61)\end{aligned}$$

This is a consequence of Lemma 10, as shown hereafter in three steps.

Step one: preliminary. We will use the following notation: for all $\beta \in B_n$ and $\epsilon \in \mathcal{E}$,

$$\begin{aligned}\Delta Q_{Y,\beta}(\epsilon) &\equiv f_{r(Q_{Y,\beta})}(Q_{Y,\beta}, G_n(q_{Y,\beta}), \epsilon) - f_{r(Q_{Y,\beta})}(Q_{Y,\beta_0}, g_0, \epsilon_0(r(Q_{Y,\beta}))) \quad \text{and} \\ \Delta d_{Y,\beta}(\epsilon) &\equiv f'_{r(Q_{Y,\beta})}(Q_{Y,\beta}, G_n(q_{Y,\beta}), \epsilon) - f'_{r(Q_{Y,\beta})}(Q_{Y,\beta_0}, g_0, \epsilon_0(r(Q_{Y,\beta}))),\end{aligned}$$

where $f_\rho(Q_Y, g, \epsilon)$ and $f'_\rho(Q_Y, g, \epsilon)$ are respectively characterized in (50) and by

$$f'_\rho(Q_Y, g, \epsilon)(O, Z) \equiv \frac{\mathbf{1}\{A = \rho(W)\}}{Z}(Y - f_\rho(Q_Y, g, \epsilon)(O)). \tag{62}$$

The two next steps mainly consist in controlling the uniform entropy integrals of the two following sets:

$$\begin{aligned}
\mathcal{F}_n &\equiv \{\Delta d_{Y,\beta}(\epsilon) : \beta \in B_n, \epsilon \in \mathcal{E}\}, \\
\mathcal{F}'_n &\equiv \{\Delta Q_{Y,\beta}(\epsilon) \circ r(Q_{Y,\beta}) : \beta \in B_n, \epsilon \in \mathcal{E}\}.
\end{aligned}$$

From now on, we assume that n is taken large enough to ensure $\beta_0 \in B_n$ and $G_n = G_\infty$, $\Delta d_{Y,\beta_0}(\epsilon_0(r_0)) = \Delta Q_{Y,\beta_0}(\epsilon_0(r_0)) = 0$ (recall that $r_0 \equiv r(Q_{Y,\beta_0})$). Consequently, $0 \in \mathcal{F}_n$ and $0 \in \mathcal{F}'_n$.

Step two: studying the first RHS term in (61). Since Z is bounded away from 0 and 1, the elements of $\mathcal{F}_n$ are uniformly bounded by a constant c which can serve as an envelope function for $\mathcal{F}_n$. We assume without loss of generality that $c \geq \max(1, \sup_{\epsilon\in\mathcal{E}} |\epsilon|)$. Obviously, ***(a)*** in Lemma 10 is met for $\mathcal{F}_n$ by the resulting (constant) sequence of (constant) envelope functions. Moreover, $\Delta d_{Y,\beta_n}(\epsilon_n) - \Delta d_{Y,\beta_0}(\epsilon_0(r_0)) = \Delta d_{Y,\beta_n}(\epsilon_n) = d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n}$ satisfies

$$\begin{aligned}
&|(\Delta d_{Y,\beta_n}(\epsilon_n) - \Delta d_{Y,\beta_0}(\epsilon_0(r_0)))(O, Z)| \\
&\qquad = \left| \frac{\mathbf{1}\{A = r_n(W)\}}{Z} \left(Q^*_{Y,\zeta_n,r_n}(A, W) - Q^*_{Y,\zeta_0,r_n}(A, W)\right) \right| \\
&\qquad \lesssim \left| Q^*_{Y,\zeta_n,r_n}(A, W) - Q^*_{Y,\zeta_0,r_n}(A, W) \right|
\end{aligned}$$

hence the convergence in probability $\|\Delta d_{Y,\beta_n}(\epsilon_n) - \Delta d_{Y,\beta_0}(\epsilon_0(r_0))\|_{2,P_{Q_0},g^{\mathrm{ref}}} = o_P(1)$ follows from the second step of the proof of Proposition 7, whose assumptions are met.

It remains to prove that $\mathcal{F}_n$ is separable and satisfies ***(b)*** in Lemma 10. Set arbitrarily $(\beta_1, \epsilon_1), (\beta_2, \epsilon_2) \in B_n \times \mathcal{E}$ and define $g_1 \equiv G_n(q_{Y,\beta_1})$, $g_2 \equiv G_n(q_{Y,\beta_2})$, $\rho_1 \equiv r(Q_{Y,\beta_1})$ and $\rho_2 \equiv r(Q_{Y,\beta_2})$. First, we note that

$$\begin{aligned}
&|(\Delta d_{Y,\beta_1}(\epsilon_1) - \Delta d_{Y,\beta_2}(\epsilon_2))(O, Z)| \\
&\qquad = \left| \frac{\mathbf{1}\{A = \rho_1(W)\}}{Z} \Delta Q_{Y,\beta_1}(\epsilon_1)(O) - \frac{\mathbf{1}\{A = \rho_2(W)\}}{Z} \Delta Q_{Y,\beta_2}(\epsilon_2)(O) \right| \\
&\qquad \lesssim |\mathbf{1}\{A = \rho_1(W)\}(\Delta Q_{Y,\beta_1}(\epsilon_1)(O) - \Delta Q_{Y,\beta_2}(\epsilon_2)(O)) \\
&\qquad\quad + |(\mathbf{1}\{A = \rho_1(W)\} - \mathbf{1}\{A = \rho_2(W)\})\Delta Q_{Y,\beta_2}(\epsilon_2)(O)|,
\end{aligned}$$

which yields the pointwise inequality

$$|\Delta d_{Y,\beta_1}(\epsilon_1) - \Delta d_{Y,\beta_2}(\epsilon_2)| \lesssim |\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)| + |\rho_1 - \rho_2|. \tag{63}$$

Second, we focus on the left RHS term in (63). It holds pointwise that

$$\begin{aligned}
|\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)| \leq\ & |f_{\rho_1}(Q_{Y,\beta_1}, g_1, \epsilon_1) - f_{\rho_1}(Q_{Y,\beta_2}, g_2, \epsilon_2)| \\
& + |f_{\rho_1}(Q_{Y,\beta_0}, g_0, \epsilon_0(\rho_1)) - f_{\rho_1}(Q_{Y,\beta_0}, g_0, \epsilon_0(\rho_2))| \\
& + |f_{\rho_1}(Q_{Y,\beta_2}, g_2, \epsilon_2) - f_{\rho_2}(Q_{Y,\beta_2}, g_2, \epsilon_2)| \\
& + |f_{\rho_1}(Q_{Y,\beta_0}, g_0, \epsilon_0(\rho_2)) - f_{\rho_2}(Q_{Y,\beta_0}, g_0, \epsilon_0(\rho_2))|.
\end{aligned}$$

For the same reasons as those which lead to (51) and because G_n c_∞-Lipschitz implies $|g_1 - g_2| \lesssim |q_{Y,\beta_1} - q_{Y,\beta_2}| \leq |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}|$, we derive the following pointwise inequalities from the previous one:

$$\begin{aligned}
|\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)| &\lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |g_1 - g_2| + |\epsilon_1 - \epsilon_2| + |\epsilon_0(\rho_1) - \epsilon_0(\rho_2)| \\
&\quad + |H_{\rho_1}(g_2) - H_{\rho_2}(g_2)| + |H_{\rho_1}(g_0) - H_{\rho_2}(g_0)| \\
&\lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| \\
&\quad + |\epsilon_1 - \epsilon_2| + |\epsilon_0(\rho_1) - \epsilon_0(\rho_2)| \\
&\quad + |H_{\rho_1}(g_2) - H_{\rho_2}(g_2)| + |H_{\rho_1}(g_0) - H_{\rho_2}(g_0)|. \qquad (64)
\end{aligned}$$

Consider the last term in the above RHS sum. Because $\mathcal{G}_1$ is uniformly bounded away from 0 and 1, we have $|H_{\rho_1}(g_0)(O) - H_{\rho_2}(g_0)(O)| \lesssim |\mathbf{1}\{A = \rho_1(W) - \mathbf{1}\{A = \rho_2(W)\}| = |\rho_1(W) - \rho_2(W)|$ (we already used this argument to derive (47) in the first step of the proof of Proposition 7). The last but one term is dealt with similarly. It remains to control the most delicate term, $|\epsilon_0(\rho_1) - \epsilon_0(\rho_2)|$. Let $\mathcal{Z}_1, \mathcal{Z}_2$ be characterized over $\mathcal{E}$ by $\mathcal{Z}_j(\epsilon) \equiv P_{Q_0,g_0} D_{Y,\rho_j}(Q_{Y,\zeta_0,\rho_j}(\epsilon), g_0) = P_{Q_0,g_0} f_{\rho_j,\epsilon}$ ($j = 1, 2$; see (45) for the definition of $f_{\rho,\epsilon}$). For the same reasons as in the first step of the proof of Proposition 7 (substitute ρ_1 or ρ_2 for r_n), $\mathcal{Z}_1(\epsilon_0(\rho_1)) = \mathcal{Z}_2(\epsilon_0(\rho_2)) = 0$ and $|\epsilon - \epsilon_0(\rho_2)| \lesssim |\mathcal{Z}_2(\epsilon)|$ for all $\epsilon \in \mathcal{E}$. Moreover, by (47), $|\mathcal{Z}_1(\epsilon) - \mathcal{Z}_2(\epsilon)| \lesssim \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}}$ for all $\epsilon \in \mathcal{E}$, hence in particular $|\mathcal{Z}_2(\epsilon_0(\rho_1))| \lesssim \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}}$. This entails the bound

$$|\epsilon_0(\rho_1) - \epsilon_0(\rho_2)| \lesssim \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}}. \qquad (65)$$

Consequently, (64) implies the pointwise inequality

$$\begin{aligned}
|\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)| &\lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| \\
&\quad + |\rho_1 - \rho_2| + \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2|. \qquad (66)
\end{aligned}$$

Combining (63) and (66) finally yields (with the same notation as in (54))

$$\begin{aligned}
|\Delta d_{Y,\beta_1}(\epsilon_1) - \Delta d_{Y,\beta_2}(\epsilon_2)| &\lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| \\
&\quad + |\rho_1 - \rho_2| + \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2|. \qquad (67)
\end{aligned}$$

Inequality (67) entails that $\mathcal{F}_n$ is separable because $\mathcal{Q}_{1,n}$, $r(\mathcal{Q}_{1,n})$ and $\mathcal{E}$ (seen as a class of constant functions with constant envelope function c') are separable. Moreover, since the definition of the *uniform* entropy integral involve a *supremum* over probability measures, (49) and (67) also imply that, for each $\delta > 0$,

$$J_c(\delta, \mathcal{F}_n) \lesssim J_c(\delta, \mathcal{Q}_{1,n}) + J_c(\delta, r(\mathcal{G}_{1,n})) + J_c(\delta, \mathcal{E}).$$

Consequently, **A4*** guarantees that ***(b)*** in Lemma 10 is met. Thus, Lemma 10 applies and yields $\sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(d^*_{Y,\zeta_n,r_n} - d^*_{Y,\zeta_0,r_n}) = \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})\Delta d_{Y,\beta_n}(\epsilon_n) = o_P(1)$.

Step three: studying the second RHS term in (61). The elements of $\mathcal{F}'_n$ are uniformly bounded by 1 hence by a constant $c' \geq \max(1, \sup_{\epsilon\in\mathcal{E}} |\epsilon|)$ which can serve as an envelope function for $\mathcal{F}'_n$. Obviously, ***(a)*** in Lemma 10 is met for $\mathcal{F}'_n$ by the resulting (constant) sequence of (constant) envelope functions. Moreover, (54) implies that $\Delta Q_{Y,\beta_n}(\epsilon_n) \circ r(Q_{Y,\beta_n}) - \Delta Q_{Y,\beta_0}(\epsilon_0(r_0)) \circ r(Q_{Y,\beta_0}) = \Delta Q_{Y,\beta_n}(\epsilon_n) \circ r_n = (Q_{Y,\zeta_n,r_n}(\epsilon_n) - Q_{Y,\zeta_0,r_n}(\varepsilon_0(r_n))) \circ r_n = (Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n$ satisfies (with the same notational convention)

$$
\begin{aligned}
&|\Delta Q_{Y,\beta_n}(\epsilon_n) \circ r(Q_{Y,\beta_n}) - \Delta Q_{Y,\beta_0}(\epsilon_0(r_0)) \circ r(Q_{Y,\beta_0})| \\
&\qquad \lesssim |Q_{Y,\beta_n} - Q_{Y,\beta_0}| + |Q^-_{Y,\beta_n} - Q^-_{Y,\beta_0}| + |g_n - g_0| + |\epsilon_n - \epsilon_0(r_n)|.
\end{aligned}
$$

Because g^{ref} is bounded away from 0 and 1, this yields

$$
\begin{aligned}
&\|\Delta Q_{Y,\beta_n}(\epsilon_n) \circ r(Q_{Y,\beta_n}) - \Delta Q_{Y,\beta_0}(\epsilon_0(r_0)) \circ r(Q_{Y,\beta_0})\|_{2,P_{Q_0,g^{\mathrm{ref}}}} \\
&\qquad \lesssim \|Q_{Y,\beta_n} - Q_{Y,\beta_0}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + \|g_n - g_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_n - \epsilon_0(r_n)| = o_P(1)
\end{aligned}
$$

because each term in the above RHS sum is $o_P(1)$ by Proposition 1, the first step of the proof of Proposition 7 and (49).

It remains to prove that $\mathcal{F}'_n$ is separable and satisfies ***(b)*** in Lemma 10. For this, set arbitrarily $(\beta_1, \epsilon_1), (\beta_2, \epsilon_2) \in B_n \times \mathcal{E}$, define $g_1 \equiv G_n(q_{Y,\beta_1})$, $g_2 \equiv G_n(q_{Y,\beta_2})$, $\rho_1 \equiv r(Q_{Y,\beta_1})$ and $\rho_2 \equiv r(Q_{Y,\beta_2})$, then note that

$$
\begin{aligned}
&|\Delta Q_{Y,\beta_1}(\epsilon_1) \circ \rho_1 - \Delta Q_{Y,\beta_2}(\epsilon_2) \circ \rho_2| \\
&\qquad \leq |\Delta Q_{Y,\beta_1}(\epsilon_1) \circ \rho_1 - \Delta Q_{Y,\beta_1}(\epsilon_1) \circ \rho_2| + |(\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)) \circ \rho_2|.
\end{aligned}
$$

Consider the first term in the above RHS sum. Because *(i)* it equals zero when ρ_1 and ρ_2 coincide, *(ii)* $|\rho_1 - \rho_2| \in \{0, 1\}$, and *(iii)* $|\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_1}(\epsilon_1)| \leq 2$, we see that it is actually upper-bounded by $2|\rho_1 - \rho_2|$. We now turn to the second term. By (66), it satisfies the following pointwise inequalities:

$$
\begin{aligned}
|(\Delta Q_{Y,\beta_1}(\epsilon_1) - \Delta Q_{Y,\beta_2}(\epsilon_2)) \circ \rho_2| &\lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| \\
&\qquad + |\rho_1 - \rho_2| + \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2|.
\end{aligned}
$$

We thus have

$$
\begin{aligned}
&|\Delta Q_{Y,\beta_1}(\epsilon_1) \circ \rho_1 - \Delta Q_{Y,\beta_2}(\epsilon_2) \circ \rho_2| \\
&\qquad \lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| + |\rho_1 - \rho_2| + \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\epsilon_1 - \epsilon_2|.
\end{aligned}
$$

As argued in the previous step, the above pointwise inequality yields that $\mathcal{F}'_n$ is separable and that, for each $\delta > 0$,

$$
J_{c'}(\delta, \mathcal{F}'_n) \lesssim J_{c'}(\delta, \mathcal{Q}_{1,n}) + J_{c'}(\delta, r(\mathcal{G}_{1,n})) + J_{c'}(\delta, \mathcal{E}).
$$

Consequently, **A4*** guarantees that ***(b)*** in Lemma 10 is met. Thus, Lemma 10 applies and implies $\sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n = \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})\Delta Q_{Y,\beta_n}(\epsilon_n) \circ r(Q_{Y,\beta_n}) = o_P(1)$. Combining the conclusions of steps two and three shows that (61) holds, and therefore completes the proof. $\square$

Proposition 8 has the following corollary. Proving it will complete the proof of Theorem 2.

Corollary 2 (second asymptotic linear expansion of ψ^*_n and resulting central limit theorem)**.** *Suppose that* **A1***,* **A2***,* **A3***,* **A4***,* **A4****, and* **A5** *are met. Then (31) holds. Moreover,* $\Sigma_n = \Sigma_0 + o_P(1)$ *with* $\Sigma_0 > 0$ *and* $\sqrt{n/\Sigma_n}(\psi^*_n - \psi_{0,r_n})$ *converges in law to the standard normal distribution.*

Proof of Corollary 2. This is a four-part proof.

Step one: preliminary. Recall (25), (26), (28), (29), (37), (38) and set

$$\begin{aligned} f_0 &\equiv d^*_{W,0} + E_{Q_{W,0}}(Q^*_{Y,\zeta_0,r_0} \circ r_0(W)) + d^*_{Y,0} = Q^*_{Y,\zeta_0,r_0} \circ r_0 + d^*_{Y,0}, \\ f_n &\equiv d^*_{W,n} + \psi^*_n + d^*_{Y,n} = Q^*_{Y,\zeta_n,r_n} \circ r_n + d^*_{Y,n}, \\ f_{0,n} &\equiv Q^*_{Y,\zeta_0,r_n} \circ r_n + d^*_{Y,\zeta_0,r_n}. \end{aligned}$$

A straightforward adaptation of the argument leading to (52) in step three of the proof of Proposition 7 also yields $E_{Q_{W,0}}(Q^*_{Y,\zeta_0,r_0} \circ r_0(W)) = \psi_0$. It is then apparent that $P_n(f_n - \psi^*_n) = P_{Q_0,g_0}(f_0 - \psi_0) = 0$. Now, note that Σ_0, Σ_n defined in (27) and (30) can be rewritten

$$\begin{aligned} \Sigma_0 &= P_{Q_0,g_0}(f_0 - \psi_0)^2 = P_{Q_0,g_0} f_0^2 - \psi_0^2, \\ \Sigma_n &= P_n(f_n - \psi^*_n)^2 = P_n f_n^2 - \psi^{*2}_n, \end{aligned}$$

and that $\Sigma_0 > 0$ by **A1**. Introduce also $S_n \equiv P_{Q_0,\mathbf{g}_n}(f_0 - \psi_0)^2$.

For each $(f, \zeta, r, \psi) \in \{(f_0, \zeta_0, r_0, 0), (f_0, \zeta_0, r_0, \psi_0), (f_n, \zeta_n, r_n, 0), (f_n, \zeta_n, r_n, \psi^*_n)\}$, it holds that

$$\begin{aligned} P_{Q_0,\mathbf{g}_n}(f - \psi)^2 &= \frac{1}{n}\sum_{i=1}^n P_{Q_0,g_i}(f - \psi)^2 \\ &= P_{Q_0,g_0}\left((Q^*_{Y,\zeta,r} \circ r - \psi)^2 + 2(Q^*_{Y,\zeta,r} \circ r - \psi) D_{Y,r}(Q^*_{\zeta,r}, g_0)\right) \\ &\quad + \frac{1}{n}\sum_{i=1}^n P_{Q_0,g_0} \frac{\mathbf{1}\{A = r(W)\}}{g_0 g_i}(Y - Q^*_{Y,\zeta,r})^2 \\ &= P_{Q_0,g_0}\left((Q^*_{Y,\zeta,r} \circ r - \psi)^2 + 2(Q^*_{Y,\zeta,r} \circ r - \psi) D_{Y,r}(Q^*_{Y,\zeta,r}, g_0)\right) \\ &\quad + P_{Q_0,g_0} \frac{\mathbf{1}\{A = r(W)\}}{g_0}(Y - Q^*_{Y,\zeta,r})^2 \times \frac{1}{n}\sum_{i=1}^n \frac{1}{g_i} \end{aligned}$$

and, similarly,

$$\begin{aligned} P_{Q_0,g_0}(f - \psi)^2 &= P_{Q_0,g_0}\left((Q^*_{Y,\zeta,r} \circ r - \psi)^2 + 2(Q^*_{Y,\zeta,r} \circ r - \psi) D_{Y,r}(Q^*_{Y,\zeta,r}, g_0)\right) \\ &\quad + P_{Q_0,g_0} \frac{\mathbf{1}\{A = r(W)\}}{g_0}(Y - Q^*_{Y,\zeta,r})^2 \times \frac{1}{g_0}. \end{aligned}$$

Since $(Y - Q^*_{Y,\zeta,r})^2 \leq 1$ and because g_0, g^{ref} and all g_is ($i \geq 1$) are bounded away from 0 and 1, applying the Cauchy-Schwarz inequality then yields

$$\begin{aligned} &|P_{Q_0,\mathbf{g}_n}(f - \psi)^2 - P_{Q_0,g_0}(f - \psi)^2| \\ &\qquad = \left| P_{Q_0,g_0} \frac{\mathbf{1}\{A = r(W)\}}{g_0}(Y - Q^*_{Y,\zeta,r})^2 \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{g_i} - \frac{1}{g_0}\right)\right| \\ &\qquad \lesssim P_{Q_0,g_0}\left|\frac{1}{n}\sum_{i=1}^n \frac{1}{g_i} - \frac{1}{g_0}\right| \lesssim \left\|\frac{1}{n}\sum_{i=1}^n (g_i - g_0)\right\|_{2,P_{Q_0,g^{\mathrm{ref}}}}. \end{aligned} \tag{68}$$

Step two: studying Σ_n and Σ_0. By Lemma 6 (presented after this proof), (68) teaches us that $E(S_n) = \Sigma_0 + o(1)$ and $S_n = \Sigma_0 + o_P(1)$ (take $(f, \psi) = (f_0, \psi_0)$ in (68)).

Let us show now that $\Sigma_n = \Sigma_0 + o_P(1)$ by proving $\Sigma_n - S_n + (\psi^{*2}_n - \psi_0^2) = \Sigma_n - S_n + o_P(1) = o_P(1)$. We use the following decomposition:

$$\Sigma_n - S_n + (\psi^{*2}_n - \psi_0^2) = P_n f_n^2 - P_{Q_0,\mathbf{g}_n} f_0^2$$

$$
\begin{aligned}
&= (P_n - P_{Q_0,\mathbf{g}_n}) f_n^2 + P_{Q_0,\mathbf{g}_n}(f_n^2 - f_0^2) \\
&= (P_n - P_{Q_0,\mathbf{g}_n}) f_n^2 + P_{Q_0,g_0}(f_n^2 - f_0^2) + o_P(1), \qquad (69)
\end{aligned}
$$

where the last equality holds because $P_{Q_0,\mathbf{g}_n} f^2 = P_{Q_0,g_0} f^2 + o_P(1)$ for both $f = f_0$ and $f = f_n$ by (68) (take $(f, \psi) = (f_0, 0)$ and $(f, \psi) = (f_n, 0)$) and Lemma 6. Let us consider in turn the two RHS terms in (69).

- From now on, we assume that n is taken large enough to ensure $G_n = G_\infty$. For all $\beta \in B_n$ and $\epsilon \in \mathcal{E}$, let

  $$d_{Y,\beta}(\epsilon) \equiv f'_{r(Q_{Y,\beta})}(Q_{Y,\beta}, G_n(q_{Y,\beta}), \epsilon)$$

  where f'_ρ is defined in (62). Introduce

  $$\mathcal{F}_n \equiv \{Q_{Y,\zeta,\rho}(\epsilon) \circ \rho + d_{Y,\beta}(\epsilon) : \beta \in B_n, g = G_n(q_{Y,\beta}), \zeta = (\beta, g), \rho = r(Q_{Y,\beta}), \epsilon \in \mathcal{E}\}.$$

  In particular, $f_n^2 = (Q_{Y,\zeta_n,r_n}(\epsilon_n) \circ r_n + d_{Y,\beta_n}(\epsilon_n))^2 \in (\mathcal{F}_n)^2 \equiv \{f^2 : f \in \mathcal{F}_n\}$. The following upper-bound motivates the definition of $\mathcal{F}_n$:

  $$|(P_n - P_{Q_0,\mathbf{g}_n}) f_n^2| \leq \|P_n - P_{Q_0,\mathbf{g}_n}\|_{(\mathcal{F}_n)^2}.$$

  If $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{(\mathcal{F}_n)^2} = o_P(1)$ then $(P_n - P_{Q_0,\mathbf{g}_n}) f_n^2 = o_P(1)$ too. We prove the former convergence by applying Lemma 9 and Markov's inequality.

  Since $\mathcal{F}_n$ is uniformly bounded, there exists a constant $c \geq \max(1, \sup_{\epsilon \in \mathcal{E}} |\epsilon|)$ which can serve as an envelope function to both $\mathcal{F}_n$ and $(\mathcal{F}_n)^2$. Set arbitrarily (β_1, ϵ_1), $(\beta_2, \epsilon_2) \in B_n \times \mathcal{E}$, define $g_1 = G_n(q_{Y,\beta_1})$, $g_2 = G_n(q_{Y,\beta_2})$, $\zeta_1 = (\beta_1, g_1)$, $\zeta_2 = (\beta_2, g_2)$, $\rho_1 = r(Q_{Y,\beta_1})$, $\rho_2 = r(Q_{Y,\beta_2})$, and let $f_1 \equiv Q_{Y,\zeta_1,\rho_1}(\epsilon_1) \circ \rho_1 + d_{Y,\beta_1}(\epsilon_1)$, $f_2 \equiv Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_2 + d_{Y,\beta_2}(\epsilon_2)$. Because $|f_1^2 - f_2^2| \lesssim |f_1 - f_2|$, it holds that $J_c(1, (\mathcal{F}_n)^2) \lesssim J_c(1, \mathcal{F}_n)$ and the separability of $\mathcal{F}_n$ implies that of $(\mathcal{F}_n)^2$. So, we now focus on $\mathcal{F}_n$.

  Obviously, $|f_1 - f_2| \leq |Q_{Y,\zeta_1,\rho_1}(\epsilon_1) \circ \rho_1 - Q_{Y,\zeta_2,\rho_2}(\epsilon_2) \circ \rho_2| + |d_{Y,\beta_1}(\epsilon_1) - d_{Y,\beta_2}(\epsilon_2)|$. The first RHS is controlled in (54). We deal with the second one in the same spirit as in step two of the proof of Proposition 8. First,

  $$
  \begin{aligned}
  &|(d_{Y,\beta_1}(\epsilon_1) - d_{Y,\beta_2}(\epsilon_2))(O, Z)| \\
  &\quad = \left| \frac{\mathbf{1}\{A = \rho_1(W)\}}{Z}(Y - f_{\rho_1}(Q_{Y,\beta_1}, g_1, \epsilon_1)(O)) \right. \\
  &\qquad \left. - \frac{\mathbf{1}\{A = \rho_2(W)\}}{Z}(Y - f_{\rho_2}(Q_{Y,\beta_2}, g_2, \epsilon_2)(O)) \right| \\
  &\quad \lesssim |\mathbf{1}\{A = \rho_1(W)\}(f_{\rho_1}(Q_{Y,\beta_1}, g_1, \epsilon_1)(O) - f_{\rho_2}(Q_{Y,\beta_2}, g_2, \epsilon_2)(O))| \\
  &\qquad + |(\mathbf{1}\{A = \rho_1(W)\} - \mathbf{1}\{A = \rho_2(W)\})\, f_{\rho_2}(Q_{Y,\beta_2}, g_2, \epsilon_2)(O)|
  \end{aligned}
  $$

  which yields

  $$|d_{Y,\beta_1}(\epsilon_1) - d_{Y,\beta_2}(\epsilon_2)| \lesssim |f_{\rho_1}(Q_{Y,\beta_1}, g_1, \epsilon_1) - f_{\rho_2}(Q_{Y,\beta_2}, g_2, \epsilon_2)| + |\rho_1 - \rho_2|.$$

  Second, the previous pointwise inequality implies

  $$|d_{Y,\beta_1}(\epsilon_1) - d_{Y,\beta_2}(\epsilon_2)| \lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |g_1 - g_2| + |\epsilon_1 - \epsilon_2| + |\rho_1 - \rho_2|.$$

  In summary,

  $$|f_1 - f_2| \lesssim |Q_{Y,\beta_1} - Q_{Y,\beta_2}| + |Q^-_{Y,\beta_1} - Q^-_{Y,\beta_2}| + |g_1 - g_2| + |\epsilon_1 - \epsilon_2| + |\rho_1 - \rho_2|. \quad (70)$$

Since $\mathcal{Q}_{1,n}$ hence $\mathcal{G}_{1,n}$ (already proven), $r(\mathcal{Q}_{1,n})$ and $\mathcal{E}$ (seen as a class of constant functions with constant envelope c) are separable, (70) implies that $\mathcal{F}_n$ is separable. Moreover, (70) also implies

$$J_c(1,\mathcal{F}_n) \lesssim J_c(1,\mathcal{Q}_{1,n}) + J_c(1, r(\mathcal{Q}_{1,n})) + J_c(1,\mathcal{E})$$

(see the argument following (54)) which yields in turn that $J_c(1,\mathcal{F}_n) = o(\sqrt{n})$ by **A4**. Thus, $(\mathcal{F}_n)^2$ is separable, $J_c(1,(\mathcal{F}_n)^2) = o(\sqrt{n})$, Lemma 9 applies and teaches us that $E(\|P_n - P_{Q_0,\mathbf{g}_n}\|_{(\mathcal{F}_n)^2}) = o(1)$, and finally Markov's inequality implies $\|P_n - P_{Q_0,\mathbf{g}_n}\|_{(\mathcal{F}_n)^2} = o_P(1)$. This completes the study of the first term in the RHS of (69).

- To rely more easily on all the results obtained so far, we first note that

$$\begin{aligned}
|P_{Q_0,g_0}(f_n^2 - f_0^2)| &\leq |P_{Q_0,g_0}(f_n^2 - f_{0,n}^2)| + |P_{Q_0,g_0}(f_{0,n}^2 - f_0^2)| \\
&\leq P_{Q_0,g_0}|f_n^2 - f_{0,n}^2| + P_{Q_0,g_0}|f_{0,n}^2 - f_0^2| \\
&\lesssim P_{Q_0,g_0}|f_n - f_{0,n}| + P_{Q_0,g_0}|f_{0,n} - f_0| \\
&\leq \|f_n - f_{0,n}\|_{2,P_{Q_0,g^{\text{ref}}}} + \|f_{0,n} - f_0\|_{2,P_{Q_0,g^{\text{ref}}}},
\end{aligned}$$

where the last upper-bound follows from the Cauchy-Schwarz inequality and the fact that g_0 and g^{ref} are bounded away from 0 and 1. Now,

$$\|f_n - f_{0,n}\|_{2,P_{Q_0,g^{\text{ref}}}} \leq \|(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n\|_{2,P_{Q_0,g^{\text{ref}}}} + \|d^*_{Y,n} - d^*_{Y,\zeta_0,r_n}\|_{2,P_{Q_0,g^{\text{ref}}}}$$

and we already proved that $\|(Q^*_{Y,\zeta_n,r_n} - Q^*_{Y,\zeta_0,r_n}) \circ r_n\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ (see step two of the proof of Proposition 7) and $\|d^*_{Y,n} - d^*_{Y,\zeta_0,r_n}\|_{2,P_{Q_0,g^{\text{ref}}}} = \|\Delta d_{Y,\beta_n}(\epsilon_n)\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ (see step two of proof of Proposition 8). Therefore, $\|f_n - f_{0,n}\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ and it suffices to show that $\|f_{0,n} - f_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ too to obtain the desired convergence $P_{Q_0,g_0}(f_n^2 - f_0^2) = o_P(1)$.

As previously, we first note that

$$\|f_{0,n} - f_0\|_{2,P_{Q_0,g^{\text{ref}}}} \leq \|Q^*_{Y,\zeta_0,r_n} \circ r_n - Q^*_{Y,\zeta_0,r_0} \circ r_0\|_{2,P_{Q_0,g^{\text{ref}}}} + \|d^*_{Y,\zeta_0,r_n} - d^*_{Y,\zeta_0,r_0}\|_{2,P_{Q_0,g^{\text{ref}}}}.$$

By (54) and (65) in step two of the proof of Proposition 8, it holds that

$$\begin{aligned}
\|Q^*_{Y,\zeta_0,r_n} \circ r_n - Q^*_{Y,\zeta_0,r_0} \circ r_0\|_{2,P_{Q_0,g^{\text{ref}}}} &\lesssim \|\epsilon_0(r_n) - \epsilon_0(r_0)\|_{2,P_{Q_0,g^{\text{ref}}}} \\
&\quad + \|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}} \\
&\lesssim \|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}}
\end{aligned}$$

with $\|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$ by Proposition 1, whose assumptions are met. Once again, we control the last remaining term in the same spirit as in step two of the proof of Proposition 8: from the upper-bound

$$\begin{aligned}
|(d^*_{Y,\zeta_0,r_n} - d^*_{Y,\zeta_0,r_0})(O,Z)| &\lesssim |\mathbf{1}\{A = r_n(W)\}(Q^*_{Y,\zeta_0,r_n} - Q^*_{Y,\zeta_0,r_0}))(A,W)| \\
&\quad + |\mathbf{1}\{A = r_n(W)\} - \mathbf{1}\{A = r_0(W)\}| \\
&\lesssim |\epsilon_0(r_n) - \epsilon_0(r_0)| + |r_n(W) - r_0(W)| \qquad (71)
\end{aligned}$$

we deduce that

$$\|d^*_{Y,\zeta_0,r_n} - d^*_{Y,\zeta_0,r_0}\|_{2,P_{Q_0,g^{\text{ref}}}} \lesssim \|r_n - r_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1).$$

In summary, $\|f_{0,n} - f_0\|_{2,P_{Q_0,g^{\text{ref}}}} = o_P(1)$, and this completes the study of the second term in the RHS of (69).

By combining the results of the above two-step study of the RHS sum in (69) and (69) itself we finally get the stated convergence $\Sigma_n = \Sigma_0 + o_P(1)$, thus completing step two of the current proof.

Step three: deriving (31) from (55). The asymptotic linear expansion (55) rewrites as

$$\begin{aligned}\psi_n^* - \psi_{r_n,0} &= (P_n - P_{Q_0,\mathbf{g}_n})f_{0,n} + o_P(1/\sqrt{n}) \\ &= (P_n - P_{Q_0,\mathbf{g}_n})f_0 + (P_n - P_{Q_0,\mathbf{g}_n})(f_{0,n} - f_0) + o_P(1/\sqrt{n}),\end{aligned}$$

hence (31) follows from the convergence $(P_n - P_{Q_0,\mathbf{g}_n})(f_{0,n} - f_0) = o_P(1/\sqrt{n})$, which is a consequence of Lemma 10.

For each $n \geq 1$, introduce the class

$$\mathcal{F}'_n \equiv \{Q^*_{Y,\zeta_0,\rho} \circ \rho + d^*_{Y,\zeta_0,\rho} - f_0 : \rho \in r(\mathcal{Q}_{1,n})\}.$$

In particular, $f_{0,n} - f_0 \in \mathcal{F}'_n$ (take $\rho = r_n$), and we have already proven in the previous step of the current proof that $\|f_{0,n} - f_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1)$. The class $\mathcal{F}'_n$ is uniformly bounded, so there exists a constant $c' \geq 1$ which can serve as an envelope function to both $\mathcal{F}'_n$ and $r(\mathcal{Q}_{1,n})$. Obviously, the resulting (constant) sequence of (constant) envelope functions satisfies condition ***(a)*** in Lemma 10. Set arbitrarily $\rho_1, \rho_2 \in r(\mathcal{Q}_{1,n})$. We have

$$\begin{aligned}&|(Q^*_{Y,\zeta_0,\rho_1} \circ \rho_1 + d^*_{Y,\zeta_0,\rho_1} - f_0) - (Q^*_{Y,\zeta_0,\rho_1} \circ \rho_2 + d^*_{Y,\zeta_0,\rho_2} - f_0)| \\ &\qquad \leq |Q^*_{Y,\zeta_0,\rho_1} \circ \rho_1 - Q^*_{Y,\zeta_0,\rho_1} \circ \rho_2| + |d^*_{Y,\zeta_0,\rho_1} - d^*_{Y,\zeta_0,\rho_2}|.\end{aligned}$$

By (54), (65) in step two of the proof of Proposition 8 and (71) with (ρ_1, ρ_2) substituted for (r_n, r_0), this inequality yields

$$\begin{aligned}&|(Q^*_{Y,\zeta_0,\rho_1} \circ \rho_1 + d^*_{Y,\zeta_0,\rho_1} - f_0) - (Q^*_{Y,\zeta_0,\rho_1} \circ \rho_2 + d^*_{Y,\zeta_0,\rho_2} - f_0)| \\ &\qquad \lesssim |\epsilon_0(\rho_1) - \epsilon_0(\rho_2)| + |\rho_1 - \rho_2| \lesssim \|\rho_1 - \rho_2\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + |\rho_1 - \rho_2|.\end{aligned}$$

Consequently, $\mathcal{F}'_n$ is separable because $r(\mathcal{Q}_{1,n})$ is separable. Moreover, since the definition of the *uniform* entropy integral involve a *supremum* over probability measures, the above pointwise inequality entails that, for each $\delta > 0$, $J_{c'}(\delta, \mathcal{F}'_n) \lesssim J_{c'}(\delta, r(\mathcal{Q}_{1,n}))$, so that condition ***(b)*** in Lemma 10 is met by **A4***. Applying Lemma 10 then gives $(P_n - P_{Q_0,\mathbf{g}_n})(f_{0,n} - f_0) = o_P(1/\sqrt{n})$, hence the validity of (31).

Step four: deducing the limiting normal distribution from (31). We first argue that (31) implies the converges in law to the standard normal distribution of $\sqrt{n/\Sigma_0}(\psi_n^* - \psi_0)$. This is a consequence of [24, Theorem 3.3.7] because *(i)* $S_n/E(S_n) - 1 = o_P(1)$, and *(ii)* for each $\alpha > 0$, $E(P_n f_0^2 \mathbf{1}\{f_0^2 \geq \alpha^2 n E(S_n)\}) = o(E(S_n))$ trivially holds since f_0 is bounded and $E(S_n) = \Sigma_0 + o(1)$ with $\Sigma_0 > 0$. Then Slutsky's lemma and $\Sigma_n = \Sigma_0 + o_P(1)$ yield the convergence in law of $\sqrt{n/\Sigma_n}(\psi_n^* - \psi_0)$ to the same limiting distribution. This completes the proof. □

Lemma 6. *If $\|g_n - g_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1)$, then $\|n^{-1}\sum_{i=1}^n (g_i - g_0)\|_{2,P_{Q_0,g^{\mathrm{ref}}}}$ converges to 0 both in probability and in L^1.*

Proof of Lemma 6. Since $\mathcal{G}$ is uniformly bounded, $\|g_n - g_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1)$ implies $E(\|g_n - g_0\|_{2,P_{Q_0,g^{\mathrm{ref}}}}) = o(1)$. Now, by convexity then Cesaro's lemma,

$$E\left(\left\|\frac{1}{n}\sum_{i=1}^n (g_i - g_0)\right\|_{2,P_{Q_0,g^{\mathrm{ref}}}}\right) \leq \frac{1}{n}\sum_{i=1}^n E\left(\|(g_i - g_0)\|_{2,P_{Q_0,g^{\mathrm{ref}}}}\right) = o(1).$$

This convergence in L^1 implies the convergence in probability because $\mathcal{G}$ is uniformly bounded. $\square$

A.3 Proofs of Propositions 2, 4, 5 and 6

Proof of Proposition 2. Set a probability measure $\tilde{\mu}$ on the measured space $\mathcal{A}\times\mathcal{W}$. Denote $\bar{\mu}$ the marginal probability measure induced by $\tilde{\mu}$ on $\mathcal{W}$. Let $\{\delta_n\}_{n\geq 1}$ be a sequence of positive numbers such that $\delta_n = o(1)$ and set $\varepsilon > 0$.

Since $r(\mathcal{Q}_{1,n})$ is a subset of a fixed VC-class of functions taking values in $[0,1]$, there exists a constant $c>0$ such that, for all $0<\varepsilon<1$,

$$\log \sup_{\mu} N(\varepsilon\|1\|_{2,\mu}, r(\mathcal{Q}_{1,n}), \|\cdot\|_{2,\mu}) \lesssim \log(\varepsilon^{-1}) + c$$

[28, Theorem 2.6.7], where 1 serves as a fixed (and constant) envelope function for $r(\mathcal{Q}_{1,n})$ and the supremum is taken over all probability measures μ on $\mathcal{W}$. It follows easily that $J_1(\delta_n, r(\mathcal{Q}_{1,n})) \lesssim \int_0^{\delta_n}\sqrt{\log(\varepsilon^{-1})+c}d\varepsilon = o(1)$. In particular, the choice $\delta_n = 1/\sqrt{n}$ yields $J_1(1, r(\mathcal{Q}_{1,n})) = o(\sqrt{n})$.

We now turn to $\mathcal{Q}_{1,n}$. Let $\{f_j^- : 1 \leq j \leq N(\varepsilon, \mathcal{F}^-, \|\cdot\|_{2,\bar{\mu}})\}$ and $\{f_j^+ : 1 \leq j \leq N(\varepsilon, \mathcal{F}^+, \|\cdot\|_{2,\bar{\mu}})\}$ be two collections of functions from $\mathcal{W}$ to $\mathbb{R}$ such that the unions of the $L^2(\bar{\mu})$-balls of radius ε centered at f_j^- or f_j^+ cover $\mathcal{F}^-$ or $\mathcal{F}^+$, respectively. Choose arbitrarily $Q_{Y,\beta} \in \mathcal{Q}_{1,n}$, with $\beta \equiv (f^-, f^+) \in B_n$. We may assume without loss of generality that $\|f^- - f_1^-\|_{2,\bar{\mu}} \leq \varepsilon$ and $\|f^+ - f_1^+\|_{2,\bar{\mu}} \leq \varepsilon$. Introduce $\beta_1 \equiv (f_1^-, f_1^+)$ and Q_{Y,β_1} defined as in (24) with β_1 substituted for β (the fact that β_1 may fall outside B_n is not a concern). Now, observe that

$$|Q_{Y,\beta} - Q_{Y,\beta_1}|^2 \leq \left(|f^- - f_1^-| + |f^+ - f_1^+|\right)^2 \leq 2\left(|f^- - f_1^-|^2 + |f^+ - f_1^+|^2\right)$$

hence

$$\|Q_{Y,\beta} - Q_{Y,\beta_1}\|_{2,\tilde{\mu}} \leq \sqrt{2}\left(\|f^- - f_1^-\|_{2,\bar{\mu}} + \|f^+ - f_1^+\|_{2,\bar{\mu}}\right) \leq 2\sqrt{2}\varepsilon.$$

This entails that $N(\varepsilon, \mathcal{Q}_{1,n}, \|\cdot\|_{2,\tilde{\mu}}) \leq N(\varepsilon/2\sqrt{2}, \mathcal{F}^-, \|\cdot\|_{2,\tilde{\mu}}) \times N(\varepsilon/2\sqrt{2}, \mathcal{F}^+, \|\cdot\|_{2,\tilde{\mu}})$. Since $\|1\|_{2,\bar{\mu}} = 1$, $\|2\|_{2,\bar{\mu}} = 2$ and because $\|1\|_{2,\tilde{\mu}} = 1$ where 1 serves as a (constant) envelope function to $\mathcal{Q}_{1,n}$, (22), (23) and the previous bound imply the existence of $\alpha \in [0,1)$ (independent of $\tilde{\mu}$) such that

$$\sqrt{\log N(\varepsilon\|1\|_{2,\tilde{\mu}}, \mathcal{Q}_{1,n}, \|\cdot\|_{2,\tilde{\mu}})} \lesssim \left(\frac{1}{\varepsilon}\right)^{\alpha}. \tag{72}$$

Taking the supremum over all probability measures $\tilde{\mu}$ on the measured space $\mathcal{A}\times\mathcal{W}$ and integrating wrt ε then yield $J_1(\delta_n, \mathcal{Q}_{1,n}) = o(1)$. In particular, the choice $\delta_n = 1/\sqrt{n}$ gives $J_1(1, \mathcal{Q}_{1,n}) = o(\sqrt{n})$.

We now turn to $L^{\mathrm{ls}}(\mathcal{Q}_{1,n})$, which admits 1 as a (constant) envelope function. Simply observe that

$$\begin{aligned}
|L(Q_{Y,\beta})(O) - L(Q_{Y,\beta_1})(O)| &= |(Y - Q_{Y,\beta}(A,W))^2 - (Y - Q_{Y,\beta_1}(A,W))^2| \\
&= |2Y - Q_{Y,\beta}(O) - Q_{Y,\beta_1}(O)| \times |Q_{Y,\beta}(O) - Q_{Y,\beta_1}(O)| \\
&\lesssim |Q_{Y,\beta}(O) - Q_{Y,\beta_1}(O)|,
\end{aligned}$$

which entails $J_1(1, L^{\mathrm{ls}}(\mathcal{Q}_{1,n})) = O(J_1(1, \mathcal{Q}_{1,n})) = o(1)$. This completes the proof. $\square$

Proof of Proposition 4. Arbitrarily set $t > 0$. By the LHS equality in (42), shown while proving Lemma 2, we first get

$$\begin{aligned}
0 \leq \psi_0 - \psi_{r_n,0} &\leq E_{Q_0}\left(|q_{Y,0}(W)| \times |r_n(W) - r_0(W)|\right) \\
&= E_{Q_0}\left(|q_{Y,0}(W)| \times \mathbf{1}\{r_n(W) \neq r_0(W)\}\right) \\
&= E_{Q_0}\left(|q_{Y,0}(W)| \times \mathbf{1}\{r_n(W) \neq r_0(W)\}\right. \\
&\qquad \left. \times \left(\mathbf{1}\{|q_{Y,0}(W)| \geq t\} + \mathbf{1}\{|q_{Y,0}(W)| < t\}\right)\right).
\end{aligned}$$

Recall that $r_n(W) \neq r_0(W)$ is equivalent to $q_{Y,\beta_n} q_{Y,\beta_0}(W) < 0$ and therefore implies $|(q_{Y,\beta_n} - q_{Y,0})(W)| \geq |q_{Y,0}(W)|$. Thus, the above inequality entails

$$\begin{aligned}
0 \leq \psi_0 - \psi_{r_n,0} \leq E_{Q_0}&\left(|(q_{Y,\beta_n} - q_{Y,0})(W)| \times \mathbf{1}\{|(q_{Y,\beta_n} - q_{Y,0})(W)| \geq |q_{Y,0}(W)| \geq t\}\right) \\
&+ E_{Q_0}\left(|q_{Y,0}(W)|^{1/3} \times |(q_{Y,\beta_n} - q_{Y,0})(W)|^{2/3} \times \mathbf{1}\{|q_{Y,0}(W)| < t\}\right). \quad (73)
\end{aligned}$$

First, we note that the left term in the above RHS expression is bounded by

$$\begin{aligned}
&E_{Q_0}\left(|(q_{Y,\beta_n} - q_{Y,0})(W)| \times \mathbf{1}\{|(q_{Y,\beta_n} - q_{Y,0})(W)| \geq |q_{Y,0}(W)| \geq t\}\right) \\
&\qquad \leq E_{Q_0}\left(\frac{|q_{Y,0}(W)|}{t} \times \frac{(q_{Y,\beta_n} - q_{Y,0})^2(W)}{t}\right) = t^{-2}\|q_{Y,\beta_n} - q_{Y,0}\|_2^2.
\end{aligned}$$

Second, Hölder's inequality and **A5*** yield that the right term in the RHS expression of (73) is bounded by

$$\|q_{Y,\beta_n} - q_{Y,0}\|_2^{2/3} \times P_{Q_0}\left(0 < |q_{Y,0}(W)| \leq t\right)^{2/3} \lesssim t^{2\gamma_2/3}\|q_{Y,\beta_n} - q_{Y,0}\|_2^{2/3}.$$

In summary, we have proven that

$$0 \leq \psi_0 - \psi_{r_n,0} \lesssim t^{-2}\|q_{Y,\beta_n} - q_{Y,0}\|_2^2 + t^{2\gamma_2/3}\|q_{Y,\beta_n} - q_{Y,0}\|_2^{2/3}.$$

Optimizing in t finally yields (32). In conclusion, $\|q_{Y,\beta_n} - q_{Y,0}\|_2 = o_P(1/n^{\gamma_3})$ does imply $\psi_0 - \psi_{r_n,0} = o_P(1/\sqrt{n})$ because $2(1+\gamma_2)/(3+\gamma_2) \times \gamma_3 = 1/2$.

The claim on the confidence interval readily follows from Proposition 3 and the property $\psi_0 - \psi_{r_n,0} = o_P(1/\sqrt{n})$. This completes the proof. □

Proof of Proposition 5. Since ψ_n^* and $n^{-1}\sum_{i=1}^n Y_i$ are known quantities, we focus on

$$\sqrt{n}\Omega_n^{\mathcal{E}} \equiv \sqrt{n}\left(\psi_n^* + \mathcal{E}_n - \frac{1}{n}\sum_{i=1}^n Y_i\right) = \sqrt{n}\left(\psi_n^* - P_n Q_{Y,0} \circ r_n\right). \quad (74)$$

By definition of $\psi_{r_n,0}$ (21) and (31), it holds that

$$\begin{aligned}
\sqrt{n}\Omega_n^{\mathcal{E}} &= \sqrt{n}(\psi_n^* - \psi_{r_n,0}) - \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})Q_{Y,0} \circ r_n \\
&= \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + d_{W,0}^* - Q_{Y,0} \circ r_0) \\
&\qquad + \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})Q_{Y,0} \circ (r_n - r_0) + o_P(1).
\end{aligned}$$

Arguments similar to those developed in Section A.2 to prove Corollary 2 successively yield $\Sigma_n^{\mathcal{E}} = \Sigma_0^{\mathcal{E}} + o_P(1)$, $\sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})Q_{Y,0} \circ (r_n - r_0) = o_P(1)$,

$$\sqrt{n}\Omega_n^{\mathcal{E}} = \sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + Q_{W,\zeta_0,r_0}^* - Q_{Y,0} \circ r_0) + o_P(1) \quad (75)$$

and the convergence in distribution of $\sqrt{n/\Sigma_n^{\mathcal{E}}}\Omega_n^{\mathcal{E}}$ to the standard normal distribution. This justifies the validity of the proposed asymptotic confidence interval. □

Proof of Proposition 6. This is a three-step proof.

Step one: preliminary. Let us assume for the time being that we also observe the variables $U_1, \ldots, U_n$ in addition to $O_1, \ldots, O_n$. The resulting counterpart to $\mathbf{O}_n$ is denoted $\mathbb{O}_n \equiv ((O_1, U_1), \ldots, (O_n, U_n))$ with convention $\mathbb{O}_0 \equiv \emptyset$. Likewise, the resulting counterpart to the empirical measure P_n is $\mathbb{P}_n$. Since the sequence $\{U_n\}_{n\geq 1}$ consists of i.i.d. variables independent from $\{O_n\}_{n\geq 1}$, a distribution $P_{Q,g} \in \mathcal{M}$ for (O, Z) yields univocally a distribution $\mathbb{P}_{Q,g}$ for (O, Z, U). For a measurable function $f : \mathcal{O} \times [0, 1] \times \mathcal{U} \to \mathbb{R}^d$, we denote $\mathbb{P}_n f \equiv n^{-1} \sum_{i=1}^n f(O_i, Z_i, U_i)$ and $\mathbb{P}_{Q,g} f \equiv E_{\mathbb{P}_{Q,g}}(f(O, Z, U))$.

Neglecting this new source of information, we carry out the exact same statistical procedure as developed and studied in Sections 2, 3, 4, 5.1 and 5.2. If we write

$$\begin{aligned}
\mathbb{P}_{Q_0,g_i} f &\equiv E_{\mathbb{P}_{Q_0,g_i}}[f(O_i, Z_i, U_i)|\mathbb{O}_{i-1}],\\
\mathbb{P}_{Q_0,\mathbf{g}_n} f &\equiv \frac{1}{n}\sum_{i=1}^n \mathbb{P}_{Q_0,g_i} f
\end{aligned}$$

for the counterparts to $P_{Q_0,g_i} f$ and $P_{Q_0,\mathbf{g}_n} f$ (each $i = 1, \ldots, n$), then (74) reads

$$\sqrt{n}\Omega_n^{\mathcal{E}} = \sqrt{n}(\psi_n^* - \mathbb{P}_n Q_{Y,0} \circ r_n) \tag{76}$$

and (75) still holds and reads

$$\sqrt{n}\Omega_n^{\mathcal{E}} = \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + Q_{W,\zeta_0,r_0}^* - Q_{Y,0} \circ r_0) + o_P(1). \tag{77}$$

Step two: inferring in the causal world. For $\rho = r_0$ and $\rho = r_n$, we set $\mathbb{Q}_{Y,0} \circ \rho(W, U) = \mathbb{Q}_{Y,0}(\rho(W), W, U)$. Since ψ_n^* and $n^{-1}\sum_{i=1}^n Y_i$ are known quantities, we focus on

$$\sqrt{n}\Omega_n^{\mathcal{C}} \equiv \sqrt{n}\left(\psi_n^* + \mathcal{C}_n - \frac{1}{n}\sum_{i=1}^n Y_i\right) = \sqrt{n}\left(\psi_n^* - \mathbb{P}_n \mathbb{Q}_{Y,0} \circ r_n\right).$$

By (76), (77), and because (35) implies $\mathbb{P}_{Q_0,\mathbf{g}_n}(\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_n = \mathbb{P}_{Q_0,g_0}(\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_n = 0$, it holds that

$$\begin{aligned}
\sqrt{n}\Omega_n^{\mathcal{C}} &= \sqrt{n}\Omega_n^{\mathcal{E}} - \sqrt{n}\mathbb{P}_n(\mathbb{Q}_{Y,0} \circ r_n - Q_{Y,0} \circ r_n)\\
&= \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + Q_{W,\zeta_0,r_0}^* - Q_{Y,0} \circ r_0)\\
&\qquad - \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_n + o_P(1)\\
&= \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(d_{Y,0}^* + Q_{W,\zeta_0,r_0}^* - Q_{Y,0} \circ r_0)\\
&\qquad - \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_0\\
&\qquad - \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})\left((\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_n - (\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_0\right) + o_P(1).
\end{aligned}$$

Define $f_0 \equiv d_{Y,0}^* + Q_{W,\zeta_0,r_0}^* - (Q_{Y,0} \circ r_0 - \psi_0)$, $\chi_0 \equiv (\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_0$, and $\Sigma_0^{\mathcal{C}} \equiv \mathbb{P}_{Q_0,g_0}(f_0 - \chi_0)^2$. Arguments similar to those developed in Section A.2 to prove Corollary 2 successively yield

$$\begin{aligned}
\sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})((\mathbb{Q}_{Y,0} - Q_{Y,0}) \circ r_n - \chi_0) &= o_P(1),\\
\sqrt{n}\Omega_n^{\mathcal{C}} &= \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{Q_0,\mathbf{g}_n})(f_0 - \chi_0) + o_P(1)
\end{aligned}$$

and the convergence in distribution of $\sqrt{n/\Sigma_0^{\mathcal{C}}}\Omega_n^{\mathcal{C}}$ to the standard normal distribution.

Step three: inferring in the real world. At this stage, there is still one issue to solve: it is not possible to infer $\Sigma_0^{\mathfrak{C}}$ because, contrary to f_0 which is a function of O, χ_0 is a function of (O,U) and we actually do not observe $U_1,\ldots,U_n$. Fortunately, it holds that

$$\Sigma_0^{\mathfrak{C}} = P_{Q_0,g_0} f_0^2 - \mathbb{P}_{Q_0,g_0}\chi_0^2 = \Sigma_0^{\mathfrak{E}} - \mathbb{P}_{Q_0,g_0}\chi_0^2 \le \Sigma_0^{\mathfrak{E}}, \tag{78}$$

the inequality justifying our claim on the proposed asymptotic confidence interval.

It only remains to prove the LHS equality in (78), which is equivalent to $\mathbb{P}_{Q_0,g_0} f_0\chi_0 = \mathbb{P}_{Q_0,g_0}\chi_0^2$. First, we note that

$$\mathbb{P}_{Q_0,g_0} f_0\chi_0 = \mathbb{P}_{Q_0,g_0}\left(Q^*_{W,\zeta_0,r_0} - (Q_{Y,0}\circ r_0 - \psi_0)\right)\chi_0 + \mathbb{P}_{Q_0,g_0} d^*_{Y,0}\chi_0.$$

By the tower rule and (35), the first RHS term in this sum equals

$$\begin{aligned}
&E_{\mathbb{P}_{Q_0,g_0}}\Big[(Q^*_{W,\zeta_0,r_0}(W) - (Q_{Y,0}\circ r_0(W) - \psi_0)) \\
&\qquad\qquad \times E_{\mathbb{P}_{Q_0,g_0}}\left(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W)\middle|W\right)\Big] = 0.
\end{aligned}$$

Thus, proving the LHS equality in (78) boils down to showing that the second term equals $\mathbb{P}_{Q_0,g_0}\chi_0^2$. By definitions of $d^*_{Y,0}$ (26) and of Y in the causal model, the tower rule and (35) imply that the second term equals

$$\begin{aligned}
&E_{\mathbb{P}_{Q_0,g_0}}\Big[\frac{\mathbf{1}\{A=r_0(W)\}}{Z}(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W)) \\
&\qquad\qquad \times E_{\mathbb{P}_{Q_0,g_0}}\left(Y - Q^*_{Y,0}(r_0(W),W)\middle|A,W,U\right)\Big] \\
&= E_{\mathbb{P}_{Q_0,g_0}}\Big[\frac{\mathbf{1}\{A=r_0(W)\}}{Z}(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W)) \\
&\qquad\qquad \times \left(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q^*_{Y,0}(r_0(W),W)\right)\Big] \\
&= E_{\mathbb{P}_{Q_0,g_0}}\Big[\frac{\mathbf{1}\{A=r_0(W)\}}{Z}(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W))^2\Big] \\
&\quad + E_{\mathbb{P}_{Q_0,g_0}}\Big[\frac{\mathbf{1}\{A=r_0(W)\}}{Z}(Q_{Y,0}(r_0(W),W) - Q^*_{Y,0}(r_0(W),W)) \\
&\qquad\qquad \times E_{\mathbb{P}_{Q_0,g_0}}\left(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W)\middle|W\right)\Big] \\
&= E_{\mathbb{P}_{Q_0,g_0}}\left[(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W))^2 E\left(\frac{\mathbf{1}\{A=r_0(W)\}}{Z}\middle|W,U\right)\right] \\
&= E_{\mathbb{P}_{Q_0,g_0}}\left[(\mathbb{Q}_{Y,0}(r_0(W),W,U) - Q_{Y,0}(r_0(W),W))^2\right] = \mathbb{P}_{Q_0,g_0}\chi_0^2.
\end{aligned}$$

This completes the proof. □

B Technical lemmas

B.1 Lemmas for M- and Z-estimation

The first lemma is a simple adaptation of [28, Corollary 3.2.3].

Lemma 7. *Let $\mathbf{M}_n$ and $\mathcal{M}_n$ be two real-valued stochastic processes indexed by a metric space (Θ, d). Consider a sequence of subsets $\Theta_n \subset \Theta$ and the following assumptions:*

(a) *For each $n \geq 1$, there exists $\tau_n \in \Theta$ such that, for all $\varepsilon > 0$,*

$$\inf_{n\geq 1} \inf\{\mathcal{M}_n(\theta) - \mathcal{M}_n(\tau_n) : \theta \in \Theta, d(\theta, \tau_n) \geq \varepsilon\} > 0.$$

(b) *For each $n \geq 1$, there exists $\tau_n^* \in \Theta_n$ such that $\mathcal{M}_n(\tau_n^*) = \inf_{\theta\in\Theta_n} \mathcal{M}_n(\theta)$. Moreover, $\mathcal{M}_n(\tau_n^*) - \mathcal{M}_n(\tau_n) = o_P(1)$.*

(c) *It holds that $\|\mathbf{M}_n - \mathcal{M}_n\|_{\Theta_n} = o_P(1)$.*

*Under **(a)**, **(b)**, and **(c)**, if $\theta_n \in \Theta_n$ satisfies $\mathbf{M}_n(\theta_n) - \mathbf{M}_n(\tau_n^*) \leq 0$ for all $n \geq 1$, then $d(\theta_n, \tau_n) = o_P(1)$.*

The corollary below will prove useful.

Lemma 8. *Let $\mathbf{Z}_n$ and $\mathcal{Z}_n$ be two real-valued stochastic processes indexed by a metric space (Θ, d). Consider the following assumptions:*

(d) *For each $n \geq 1$, there exists $\tau_n \in \Theta$ such that $\mathcal{Z}_n(\tau_n) = 0$ and, for all $\varepsilon > 0$,*

$$\inf_{n\geq 1} \inf\{|\mathcal{Z}_n(\theta)| : \theta \in \Theta, d(\theta, \tau_n) \geq \varepsilon\} > 0.$$

(e) *It holds that $\|\mathbf{Z}_n - \mathcal{Z}_n\|_{\Theta} = o_P(1)$.*

*Under **(d)** and **(e)**, if $\theta_n \in \Theta$ satisfies $\mathbf{Z}_n(\theta_n) = 0$ for all $n \geq 1$, then $d(\theta_n, \tau_n) = o_P(1)$.*

Proof of Lemma 7. Set $n \geq 1$. By ***(a)***, it holds that

$$\begin{aligned} 0 &\leq \mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n) \\ &= (\mathcal{M}_n(\theta_n) - \mathbf{M}_n(\theta_n)) + (\mathbf{M}_n(\theta_n) - \mathbf{M}_n(t_n^*)) \\ &\quad + (\mathbf{M}_n(t_n^*) - \mathcal{M}_n(t_n^*)) + (\mathcal{M}_n(t_n^*) - \mathcal{M}_n(t_n)) \,. \end{aligned}$$

The above first and third RHS terms are both upper-bounded by $\|\mathbf{M}_n - \mathcal{M}_n\|_{\Theta_n}$. The second RHS term is non-positive by definition of θ_n. The fourth RHS terms is $o_P(1)$ by ***(b)***. Thus, it actually holds that $0 \leq \mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n) \leq 2\|\mathbf{M}_n - \mathcal{M}_n\|_{\Theta_n} + o_P(1) = o_P(1)$ by ***(c)***.

Set $\varepsilon > 0$. By ***(a)***, there exists a positive random variable Δ which is independent of n and such that $d(\theta_n, t_n) \geq \varepsilon$ implies $\mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n) \geq \Delta$ or, equivalently, $\Delta^{-1}[\mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n)] \geq 1$. Now, by Slutsky's lemma [27, Lemma 2.8], $\mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n) = o_P(1)$ entails $\Delta^{-1}[\mathcal{M}_n(\theta_n) - \mathcal{M}_n(t_n)] = o_P(1)$. Therefore, we conclude that $d(\theta_n, t_n) = o_P(1)$ too. $\square$

Proof of Lemma 8. For all $n \geq 1$ and $\theta \in \Theta$, define $\Theta_n = \Theta$, $t_n^* = t_n$, $\mathbf{M}_n(\theta) = |\mathbf{Z}_n(\theta)|$ and $\mathcal{M}_n(\theta) = |\mathcal{Z}_n(\theta)|$. We note that ***(a)*** in Lemma 7 follows from ***(d)***, that ***(b)*** in Lemma 7 trivially holds, and finally that ***(c)*** in Lemma 7 is a consequence of ***(e)*** and the reverse triangle inequality. Now, for each $n \geq 1$, $\mathbf{Z}_n(\theta_n) = 0$ rewrites $\mathbf{M}_n(\theta_n) - \mathbf{M}_n(t_n^*) \leq 0$. Applying Lemma 7 yields the result. $\square$

B.2 Maximal inequalities and convergence of empirical processes

The following two results are the cornerstones of our theoretical study.

Lemma 9 (maximal inequality)**.** *Let $\mathcal{F}$ be a separable class of measurable, real-valued functions, with envelope function F. Set $n \geq 1$. It holds that*

$$E\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}}\right) \lesssim J_F(1,\mathcal{F}) \times \|F\|_{2,P_{Q_0,g^{\mathrm{ref}}}}. \tag{79}$$

Lemma 10 (convergence of empirical processes indexed by estimated functions)**.** *For each $n \geq 1$, let $\mathcal{F}_n = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in T_n\}$ be a separable class of measurable, real-valued functions, with envelope function F_n. Suppose the following holds:*

(a) *The sequence $\{F_n\}_{n\geq 1}$ satisfies the Lindeberg condition: $\|F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = O(1)$ and, for every $\delta > 0$, $\|F_n \mathbf{1}\{F_n > \delta\sqrt{n}\}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o(1)$.*

(b) *If $\delta_n = o(1)$, then it holds that $J_{F_n}(\delta_n, \mathcal{F}_n) = o(1)$.*

If $\eta_n \in T_n$ is such that $\sup_{\theta\in\Theta} \|f_{\theta,\eta_n} - f_{\theta,\eta_0}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o_P(1)$ for some $\eta_0 \in \cap_{p\geq 1} \cup_{n\geq p} T_n$, then $\sup_{\theta\in\Theta} |\sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(f_{\theta,\eta_n} - f_{\theta,\eta_0})| = o_P(1)$.

The proof of Lemma 10 notably relies on the lemma below. Its proof, a straightforward adaptation of that of [33, Lemma 12], is omitted.

Lemma 11. *For each $n \geq 1$, let $\mathcal{F}_n$ be a class of measurable, real-valued functions with envelope function F_n such that $\delta_n = o(1)$ implies $J_{F_n}(\delta_n, \mathcal{F}_n) = o(1)$. Then (i) $J_{F_n}(\delta, \mathcal{F}_n) = O(1)$ for every $\delta > 0$, and (ii) for every $\varepsilon > 0$, there exist $\delta > 0$ and $n_1 \geq 1$ such that $J_{F_n}(\delta, \mathcal{F}_n) \leq \varepsilon$ for all $n \geq n_1$.*

Proof of Lemmas 9 and 10. The proofs of Lemmas 9 and 10 are best presented jointly.

Let us prove (79) from Lemma 9 in three steps.

Step one: decoupling. By [8, Proposition 6.1.5 and Remark 6.1.6], it is possible to enlarge the probability space and to define three sequences of random variables $\{\varepsilon_n\}_{n\geq 1}$, $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$, $\{(O_n^\natural, Z_n^\natural)\}_{n\geq 1}$ and a σ-field $\mathcal{G}$ such that

- $\{\varepsilon_n\}_{n\geq 1}$ is a sequence of independent Rademacher random variables, a sequence that is moreover independent of $\{(O_n, Z_n)\}_{n\geq 1}$, $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$, $\{(O_n^\natural, Z_n^\natural)\}_{n\geq 1}$;
- the distributions of $(O_1^\flat, Z_1^\flat)$ and $(O_1^\natural, Z_1^\natural)$ coincide with that of (O_1, Z_1) and, for every $n \geq 2$, the conditional distributions of $(O_n^\flat, Z_n^\flat)$ and $(O_n^\natural, Z_n^\natural)$ given $\mathcal{G}$ coincide with that of (O_n, Z_n) given $\{(O_1, Z_1), \ldots, (O_{n-1}, Z_{n-1})\}$;
- conditionally on $\mathcal{G}$, the two sequences $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$, $\{(O_n^\natural, Z_n^\natural)\}_{n\geq 1}$ are independent and each with mutually independent elements.

The new sequences $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$ and $\{(O_n^\natural, Z_n^\natural)\}_{n\geq 1}$ are said "decoupled sequences" to $\{(O_n, Z_n)\}_{n\geq 1}$.

We denote $E_{\mathcal{G}}$ the conditional expectation given $\mathcal{G}$ and $E_{\mathcal{G}}^\flat$ the conditional expectation given $\mathcal{G}$ and $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$. We also characterize $P_n^\flat$, $P_{Q_0,\mathbf{g}_n}^\flat$ and $P_n^{0\flat}$ by setting, for each $f : \mathcal{O} \times [0,1] \to \mathbb{R}$, $P_n^\flat f = n^{-1}\sum_{i=1}^n f(O_i^\flat, Z_i^\flat)$, $P_{Q_0,\mathbf{g}_n}^\flat f = n^{-1}\sum_{i=1}^n E_{\mathcal{G}}[f(O_i^\flat, Z_i^\flat)]$, $P_n^{0\flat} f = n^{-1}\sum_{i=1}^n \varepsilon_i f(O_i^\flat, Z_i^\flat)$.

Step two: symmetrizing. Let Φ be a non-decreasing, convex function mapping $\mathbb{R}_+$ to $\mathbb{R}$. Set $n \geq 1$. By construction of the decoupled sequences, it holds that $E[\Phi(n\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}})] = E(E_{\mathcal{G}}[\Phi(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}})])$. We now focus on $E_{\mathcal{G}}[\Phi(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}})]$.

Note that

$$\begin{aligned} n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}} &= \left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - E_{\mathcal{G}}(f(O_i^\natural, Z_i^\natural)) \right\|_{\mathcal{F}} \\ &= \left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - E_{\mathcal{G}}^\flat(f(O_i^\natural, Z_i^\natural)) \right\|_{\mathcal{F}} \\ &\leq E_{\mathcal{G}}^\flat \left[\left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural) \right\|_{\mathcal{F}} \right], \end{aligned}$$

so that Jensen's inequality yields

$$\Phi\left(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}}\right) \leq E_{\mathcal{G}}^\flat \left[\Phi\left(\left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural) \right\|_{\mathcal{F}} \right) \right].$$

By taking outer (conditional) expectation, we obtain

$$E_{\mathcal{G}}\left[\Phi\left(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}}\right)\right] \leq E_{\mathcal{G}} \left[\Phi\left(\left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural) \right\|_{\mathcal{F}} \right) \right]. \tag{80}$$

Observe now that, for every n-tuple $(e_1, \ldots, e_n) \in \{-1, 1\}^n$,

$$E_{\mathcal{G}} \left[\Phi\left(\left\| \sum_{i=1}^n f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural) \right\|_{\mathcal{F}} \right) \right] = E_{\mathcal{G}} \left[\Phi\left(\left\| \sum_{i=1}^n e_i(f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural)) \right\|_{\mathcal{F}} \right) \right]$$

since, for each $1 \leq i \leq n$, $(O_i^\flat, Z_i^\flat)$ and $(O_i^\natural, Z_i^\natural)$ are independent and equal in law (conditional on $\mathcal{G}$). Consequently, (80) yields

$$E_{\mathcal{G}}\left[\Phi\left(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}}\right)\right] \leq E_{\mathcal{G}} \left[\Phi\left(\left\| \sum_{i=1}^n \varepsilon_i(f(O_i^\flat, Z_i^\flat) - f(O_i^\natural, Z_i^\natural)) \right\|_{\mathcal{F}} \right) \right], \tag{81}$$

where the expectation $E_{\mathcal{G}}$ to the right now also concerns the (conditionally and unconditionally on $\mathcal{G}$) independent $(\varepsilon_1, \ldots, \varepsilon_n)$. By the triangle inequality and convexity of Φ, we see that the RHS expression of (81) is itself upper-bounded by

$$\begin{aligned} \frac{1}{2} E_{\mathcal{G}} \left[\Phi\left(2\left\| \sum_{i=1}^n \varepsilon_i f(O_i^\flat, Z_i^\flat) \right\|_{\mathcal{F}} \right) \right] &+ \frac{1}{2} E_{\mathcal{G}} \left[\Phi\left(2\left\| \sum_{i=1}^n \varepsilon_i f(O_i^\natural, Z_i^\natural) \right\|_{\mathcal{F}} \right) \right] \\ &= E_{\mathcal{G}} \left[\Phi\left(2\left\| \sum_{i=1}^n \varepsilon_i f(O_i^\flat, Z_i^\flat) \right\|_{\mathcal{F}} \right) \right], \end{aligned}$$

hence

$$E_{\mathcal{G}}\left[\Phi\left(n\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}}\right)\right] \leq E_{\mathcal{G}} \left[\Phi\left(2\left\| \sum_{i=1}^n \varepsilon_i f(O_i^\flat, Z_i^\flat) \right\|_{\mathcal{F}} \right) \right].$$

In conclusion, we derive the symmetrization inequality

$$E\left[\Phi\left(n\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}}\right)\right] \le E[\Phi(2n\|P_n^{0\flat}\|_{\mathcal{F}})]. \tag{82}$$

Step three: chaining. Taking Φ given by $\Phi(x) = x$ (all $x \ge 0$) in (82) readily yields

$$E\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}}\right) \le 2E(\sqrt{n}\|P_n^{0\flat}\|_{\mathcal{F}}). \tag{83}$$

Set now $\Phi(x) = \exp(x^2) - 1$ (all $x \ge 0$) and let $\|\cdot\|_\Phi$ be the corresponding Φ-Orlicz norm [28, page 95]. Conditionally on $(O_1^\flat, Z_1^\flat), \ldots, (O_n^\flat, Z_n^\flat)$, the process $\sqrt{n}P_n^{0\flat}$ is sub-Gaussian for the $L^2(P_n^\flat)$-seminorm $\|\cdot\|_{2,n}^\flat$ by Hoeffding's inequality [28, Lemma 2.2.7]. The number $s_n^\flat = \sup_{f\in\mathcal{F}} \|f\|_{2,n}^\flat$ upper-bounds the radius of $\mathcal{F} \cup \{0\}$ wrt $\|\cdot\|_{2,n}^\flat$. Thus, by [28, Theorem 2.2.4] (a maximal inequality whose proof essentially relies on a chaining argument) and a change of variable, it holds that

$$\begin{aligned}
\|\sqrt{n}P_n^{0\flat}\|_\Phi &\lesssim \int_0^{s_n^\flat} \sqrt{1 + \log N(\varepsilon, \mathcal{F}, L^2(P_n^\flat))}\,d\varepsilon \\
&\le \|F\|_{2,n}^\flat \int_0^{s_n^\flat/\|F\|_{2,n}^\flat} \sqrt{1 + \log N(\varepsilon\|F\|_{2,n}^\flat, \mathcal{F}, L^2(P_n^\flat))}\,d\varepsilon.
\end{aligned}$$

By definition of the uniform entropy integral, we therefore obtain

$$\|\sqrt{n}P_n^{0\flat}\|_\Phi \lesssim \|F\|_{2,n}^\flat J_{F_n}(s_n^\flat/\|F\|_{2,n}^\flat, \mathcal{F}),$$

a result which holds conditionally on $(O_1^\flat, Z_1^\flat), \ldots, (O_n^\flat, Z_n^\flat)$. Finally, we take the expectation wrt to $(O_1^\flat, Z_1^\flat), \ldots, (O_n^\flat, Z_n^\flat)$ and note that *(a)* $s_n^\flat \le \|F\|_{2,n}^\flat$, *(b)* $E(\|F\|_{2,n}^\flat) \lesssim \|F\|_{2,P_{Q_0,g^{\mathrm{ref}}}}$. In view of (83) this does yield

$$\begin{aligned}
E\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}}\right) &= E\left(\sqrt{n}\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}}\right) \\
&\lesssim E\left(\|F\|_{2,n}^\flat \times J_{F_n}(s_n^\flat/\|F\|_{2,n}^\flat, \mathcal{F})\right) \\
&\le J_{F_n}(1, \mathcal{F}) \times \|F\|_{2,P_{Q_0,g^{\mathrm{ref}}}},
\end{aligned} \tag{84}$$

which completes the proof of (79).

We now show Lemma 10. The proof follows closely that of [2, Part III, Theorem 6.16]. It has four steps.

Step one: preliminary. Introduce the classes $\widetilde{\mathcal{F}}_n^0$ (random) and $\mathcal{F}_n^0$ (deterministic) given by

$$\widetilde{\mathcal{F}}_n^0 \equiv \{f_{\theta,\eta_n} - f_{\theta,\eta_0} : \theta \in \Theta\} \subset \mathcal{F}_n^0 \equiv \{f_{\theta,\eta} - f_{\theta,\eta_0} : \theta \in \Theta, \eta \in T_n\}.$$

Lemma 10 states that $\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\widetilde{\mathcal{F}}_n^0} = o_P(1)$.

For an arbitrarily fixed $\delta > 0$, define

$$\begin{aligned}
T_n^0(\delta) &\equiv \left\{\eta \in T_n : \sup_{\theta\in\Theta} \|f_{\theta,\eta} - f_{\theta,\eta_0}\|_{2,P_{Q_0,g^{\mathrm{ref}}}}^2 \le \delta^2\right\}, \\
\mathcal{F}_n^0(\delta) &\equiv \{f_{\theta,\eta} - f_{\theta,\eta_0} : \theta \in \Theta, \eta \in T_n^0(\delta)\} \subset \mathcal{F}_n^0, \\
\mathcal{F}_n^0(\delta)^2 &\equiv \{h^2 : h \in \mathcal{F}_n^0(\delta)\}, \quad \text{and}
\end{aligned}$$

$$s_n^\flat(\delta) \quad \equiv \quad \frac{\sup_{h\in\mathcal{F}_n^0(\delta)}\|h\|_{2,n}^\flat}{\|1+2F_n\|_{2,n}^\flat} = \frac{\|P_n^\flat\|_{\mathcal{F}_n^0(\delta)^2}}{\|1+2F_n\|_{2,n}^\flat}.$$

The classes $\widetilde{\mathcal{F}}_n^0$, $\mathcal{F}_n^0(\delta)$ and $\mathcal{F}_n^0$ admit $H_n \equiv 1+2F_n$ as an envelope function. Because its definition involves $P_n^\flat$, $s_n^\flat(\delta)$ is random. Moreover, $\|H_n\|_{2,n}^\flat \geq 1$ and $\sup_{h\in\mathcal{F}_n^0(\delta)}\|h\|_{2,n}^\flat \leq \|2F_n\|_{2,n}^\flat$ yield that

$$s_n^\flat(\delta) \leq \min\left(1, \sup_{h\in\mathcal{F}_n^0(\delta)}\|h\|_{2,n}^\flat\right) = \min\left(1, \|P_n^\flat\|_{\mathcal{F}_n^0(\delta)^2}\right). \tag{85}$$

By (84) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left[E\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}_n^0(\delta)}\right)\right]^2 \quad &\lesssim \quad \left[E\left(\|H_n\|_{2,n}^\flat \times J_{H_n}(s_n^\flat(\delta), \mathcal{F}_n^0(\delta))\right)\right]^2 \\ &\leq \quad E\left(\|H_n\|_{2,n}^{\flat 2}\right) \times E\left(J_{H_n}(s_n^\flat(\delta), \mathcal{F}_n^0(\delta))^2\right). \end{aligned} \tag{86}$$

Step two: studying $s_n^\flat(\delta)$. We now show that there exists an integer $n_1(\delta)$ such that $E(s_n^\flat(\delta)) \lesssim \min(1,\delta^2)$ for all $n \geq n_1(\delta)$. The proof is based on (85) and the decomposition $\mathcal{F}_n^0(\delta)^2 = \mathcal{F}_{n,1}^0(\delta)^2 \cup \mathcal{F}_{n,2}^0(\delta)^2$ for

$$\begin{aligned}\mathcal{F}_{n,1}^0(\delta)^2 \quad &\equiv \quad \left\{h^2\mathbf{1}\{2F_n \leq \rho\sqrt{n}/2\} : h \in \mathcal{F}_n^0(\delta)\right\}, \\ \mathcal{F}_{n,2}^0(\delta)^2 \quad &\equiv \quad \left\{h^2\mathbf{1}\{2F_n > \rho\sqrt{n}/2\} : h \in \mathcal{F}_n^0(\delta)\right\}\end{aligned}$$

where the constant $\rho > 0$ will be determined later.

Obviously, $\rho\sqrt{n}/2 \times 2F_n = \rho\sqrt{n}F_n$ is an envelope function for $\mathcal{F}_{n,1}^0(\delta)^2$. By (84), we thus have

$$E\left(\sqrt{n}\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_{n,1}^0(\delta)^2}\right) \lesssim J_{\rho\sqrt{n}F_n}(1, \mathcal{F}_{n,1}^0(\delta)^2) \times \|\rho\sqrt{n}F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}}. \tag{87}$$

But $J_{\rho\sqrt{n}F_n}(1,\mathcal{F}_{n,1}^0(\delta)^2)$ easily compares to $J_{F_n}(1,\mathcal{F}_n)$. Indeed, whichever are $\varepsilon > 0$, $h, h' \in \mathcal{F}_n^0(\delta)$, and m a discrete probability measure such that $0 < mF_n$, it holds that

$$m(h^2 - h'^2)^2\mathbf{1}\{2F_n \leq \rho\sqrt{n}/2\} \leq (4F_n)^2 m(h-h')^2 \leq (\rho\sqrt{n})^2 m(h-h')^2,$$

hence

$$N(\varepsilon\|\rho\sqrt{n}F_n\|_{m,2}, \mathcal{F}_{n,1}^0(\delta)^2) \leq N(\varepsilon\|F_n\|_{m,2}, \mathcal{F}_n),$$

from which we deduce that $J_{\rho\sqrt{n}F_n}(1,\mathcal{F}_{n,1}^0(\delta)^2) \leq J_{F_n}(1,\mathcal{F}_n)$. This bound and (87) yield

$$E\left(\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_{n,1}^0(\delta)^2}\right) \lesssim \rho J_{F_n}(1,\mathcal{F}_n) \times \|F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}}. \tag{88}$$

Furthermore, because *(i)* $2F_n$ is an envelope function for $\mathcal{F}_{n,2}^0(\delta)^2$ and *(ii)* the design $\mathbf{g}_n$ attached to the sequence $\{(O_n^\flat, Z_n^\flat)\}_{n\geq 1}$ is bounded away from 0 and 1, it holds that

$$E\left(\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_{n,2}^0(\delta)^2}\right) \lesssim \rho J_{F_n}(1,\mathcal{F}_n) \times \|F_n\mathbf{1}\{F_n > \rho\sqrt{n}/2\}\|_{2,P_{Q_0,g^{\mathrm{ref}}}}.$$

Since $\mathcal{F}_n^0(\delta)^2$ is the union of $\mathcal{F}_{n,1}^0(\delta)^2$ and $\mathcal{F}_{n,2}^0(\delta)^2$, the previous inequality combined with (88) then yields

$$E\left(\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_n^0(\delta)^2}\right)$$
$$\lesssim \rho J_{F_n}(1,\mathcal{F}_n) \times \|F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}} + \rho J_{F_n}(1,\mathcal{F}_n) \times \|F_n \mathbf{1}\{F_n > \rho\sqrt{n}/2\}\|_{2,P_{Q_0,g^{\mathrm{ref}}}}.$$

By ***(a)*** in Lemma 10, $\|F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = O(1)$ and $\|F_n \mathbf{1}\{F_n > \rho\sqrt{n}/2\}\|_{2,P_{Q_0,g^{\mathrm{ref}}}} = o(1)$. By Lemma 11, $J_{F_n}(1,\mathcal{F}_n) = O(1)$. Therefore, it is possible to choose $\rho > 0$ and find $n_1(\delta) \geq 1$ such that, for all $n \geq n_1(\delta)$,

$$E\left(\|P_n^\flat - P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_n^0(\delta)^2}\right) \leq \delta^2. \tag{89}$$

Now, the definition of $\mathcal{F}_n^0(\delta)$ and the above remark *(ii)* about the design $\mathbf{g}_n$ yield the additional inequality, valid for all sample size:

$$E\left(\|P_{Q_0,\mathbf{g}_n}^\flat\|_{\mathcal{F}_n^0(\delta)^2}\right) \lesssim \delta^2. \tag{90}$$

By the triangle inequality, (85), (89) and (90) imply

$$E(s_n^\flat(\delta)) \leq \min\left(1, E\left(\|P_n^\flat\|_{\mathcal{F}_n^0(\delta)^2}\right)\right) \lesssim \min(1,\delta^2)$$

for all $n \geq n_1(\delta)$. Markov's inequality then yields that, for all $\xi > 0$ and $n \geq n_1(\delta)$,

$$P\left(s_n^\flat(\delta) \geq \xi\right) \leq \xi^{-1}\min(1,\delta^2). \tag{91}$$

This completes the study of $s_n^\flat(\delta)$.

Step three: fine-tuning. Set arbitrarily $\alpha, \varepsilon > 0$. Note that the above remark *(ii)* about the design $\mathbf{g}_n$ and assumption ***(a)*** in Lemma 10 imply the existence of a constant $C_1 > 0$ such that the following bounds hold on the leftmost factor of the RHS expression in (86):

$$E\left(\|H_n\|_{2,n}^{\flat 2}\right) \lesssim \|F_n\|_{2,P_{Q_0,g^{\mathrm{ref}}}}^2 \leq C_1^2. \tag{92}$$

By assumption ***(b)*** in Lemma 10 and Lemma 11, there exist $0 < \xi \leq 1$, $C_2 > 0$ and $n_2 \geq 1$ such that $J_{H_n}(\xi,\mathcal{F}_n) \leq \alpha\varepsilon/C_1$ and $J_{H_n}(1,\mathcal{F}_n)^2 \leq C_2^2$ for all $n \geq n_2$. Let $\delta_0 > 0$ be such that $\delta_0 \leq \alpha\varepsilon\sqrt{3\xi}/C_1C_2$. By assumption on η_n in Lemma 10, we know that there exists $n_3(\delta_0) \geq 1$ such that $P(\eta_n \notin T_n^0(\delta_0)) \leq \varepsilon$ for all $n \geq n_3(\delta_0)$.

Step four: wrapping up. Let n be larger than $\max(n_1(\delta_0), n_2, n_3(\delta_0))$. It holds that

$$\begin{aligned}
A \equiv P&\left(\sup_{\theta\in\Theta}|\sqrt{n}(P_n - P_{Q_0,\mathbf{g}_n})(f_{\theta,\eta_n} - f_{\theta,\eta_0})| \geq \alpha\right)\\
&= P\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\widetilde{\mathcal{F}}_n^0} \geq \alpha\right)\\
&\leq P\left(\eta_n \notin T_n^0(\delta_0)\right) + P\left(\eta_n \in T_n^0(\delta_0),\ \sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\widetilde{\mathcal{F}}_n^0} \geq \alpha\right)\\
&\leq \varepsilon + P\left(\sqrt{n}\|P_n - P_{Q_0,\mathbf{g}_n}\|_{\mathcal{F}_n^0(\delta_0)} \geq \alpha\right).
\end{aligned}$$

By Markov's inequality, (86), (92) and (91), we obtain the inequalities

$$\begin{aligned}
A &\leq \varepsilon + \alpha^{-1}E\left(\|H_n\|_{2,n}^{\flat 2}\right)^{1/2} \times E\left(J_{H_n}(s_n^\flat(\delta), \mathcal{F}_n^0(\delta))^2\right)^{1/2}\\
&\leq \varepsilon + \alpha^{-1}C_1 \times \left(P(s_n^\flat(\delta_0) \geq \xi) \times J_{H_n}(1,\mathcal{F}_n)^2 + J_{H_n}(\xi,\mathcal{F}_n)^2\right)^{1/2}\\
&\leq \varepsilon + \alpha^{-1}C_1 \times \left(\xi^{-1}\min(1,\delta_0^2) \times C_2^2 + (C_1^{-1}\alpha\varepsilon)^2\right) \leq 3\varepsilon.
\end{aligned}$$

Since α and ε were arbitrarily chosen, this completes the proof of Lemma 10. □

C Pathwise differentiability

The next two lemmas are summaries of results stated and shown in [15, 16]. We state them for the sake of completeness.

Lemma 12. *Set $\rho \in \mathcal{R}$, a* known *treatment rule. Let $\Psi_\rho : \mathcal{M} \to [0,1]$ be given by*

$$\Psi_\rho(P_{Q,g}) \equiv E_Q\left(Q_Y(\rho(W), W)\right). \tag{93}$$

The mapping $\Psi_\rho : \mathcal{M} \to [0,1]$ is pathwise differentiable at every $P_{Q,g} \in \mathcal{M}$ with respect to (wrt) the maximal tangent space. Its efficient influence curve at $P_{Q,g}$ is $D_\rho(Q,g)$ which satisfies $D_\rho(Q,g)(O) = D_{W,\rho}(Q,g)(W) + D_{Y,\rho}(Q,g)(O)$ with

$$\begin{aligned} D_{W,\rho}(Q)(W) &\equiv Q_Y(\rho(W), W) - \Psi_\rho(P_{Q,g}), \\ D_{Y,\rho}(Q,g)(O) &\equiv \frac{\mathbf{1}\{A = \rho(W)\}}{g(A|W)}\left(Y - Q_Y(A,W)\right). \end{aligned}$$

The variance $\mathrm{Var}_{P_{Q,g}} D_\rho(Q,g)(O)$ is a generalized Cramér-Rao lower bound for the asymptotic variance of any regular and asymptotically linear estimator of $\Psi_\rho(P_{Q,g})$ when sampling independently from $P_{Q,g}$.

In addition, if $g = g'$, then $E_{Q,g}(D_\rho(Q',g')(O)) = 0$ implies $\Psi_\rho(P_{Q',g'}) = \Psi_\rho(P_{Q,g})$.

The notation $D_{W,\rho}(Q)$ conveys the notion that the first component of $D_\rho(Q,g)$ does not depend on g. This is true because $\Psi_\rho(P_{Q,g})$ does not depend on g either.

Lemma 13. *The mapping $\Psi : \mathcal{M} \to [0,1]$ is pathwise differentiable at every $P_{Q,g} \in \mathcal{M}$ wrt the maximal tangent space. Its efficient influence curve at $P_{Q,g}$ is $D(Q,g)$ which satisfies $D(Q,g)(O) = D_W(Q,g)(W) + D_Y(Q,g)(O)$ with*

$$\begin{aligned} D_W(Q)(W) &\equiv Q_Y(r(Q_Y)(W), W) - \Psi(P_{Q,g}), \\ D_Y(Q,g)(O) &\equiv \frac{\mathbf{1}\{A = r(Q_Y)(W)\}}{g(A|W)}\left(Y - Q_Y(A,W)\right). \end{aligned}$$

The variance $\mathrm{Var}_{P_{Q,g}} D(Q,g)(O)$ is a generalized Cramér-Rao lower bound for the asymptotic variance of any regular and asymptotically linear estimator of $\Psi(P_{Q,g})$ when sampling independently from $P_{Q,g}$.

In addition, if $g = g'$, then $E_{Q,g}(D(Q',g')(O)) = 0$ implies

$$\Psi(P_{Q',g'}) = E_Q\left(Q_Y(r(Q'_Y)(W), W)\right).$$

In particular, if $r(Q_Y) = r(Q'_Y)$ and $g = g'$, then $E_{Q,g}(D(P_{Q',g'})(O)) = 0$ implies $\Psi(P_{Q',g'}) = \Psi(P_{Q,g})$.

References

[1] L. B. Balzer, M. L. Petersen, and M. J. van der Laan. Targeted estimation and inference for the sample average treatment effect. Technical Report 334, U.C. Berkeley Division of Biostatistics Working Paper Series, 2015. URL `http://biostats.bepress.com/ucbbiostat/paper334`.

[2] E. Bolthausen, E. Perkins, and A. van der Vaart. *Lectures on probability theory and statistics*, volume 1781 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. doi: 10.1007/b93152. URL `http://dx.doi.org/10.1007/b93152`. Lectures from the 29th Summer School on Probability Theory held in Saint-Flour, July 8–24, 1999, Edited by P. Bernard.

[3] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1–122), 2012.

[4] B. Chakraborty and E. E. M. Moodie. *Statistical methods for dynamic treatment regimes*. Statistics for Biology and Health. Springer, New York, 2013. doi: 10.1007/978-1-4614-7428-9. URL `http://dx.doi.org/10.1007/978-1-4614-7428-9`. Reinforcement learning, causal inference, and personalized medicine.

[5] B. Chakraborty, E. B. Laber, and Y-Q. Zhao. Inference about the expected performance of a data-driven dynamic treatment regime. *Clin. Trials*, 11(4):408–417, 2014.

[6] A. Chambaz and M. J. van der Laan. Inference in targeted group-sequential covariate-adjusted randomized clinical trials. *Scand. J. Stat.*, 41(1):104–140, 2014. doi: 10.1111/sjos.12013. URL `http://dx.doi.org/10.1111/sjos.12013`.

[7] A. Chambaz, M. J. van der Laan, and W. Zheng. Targeted covariate-adjusted response-adaptive lasso-based randomized controlled trials. In A. Sverdlov, editor, *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects*, pages 345–368. CRC Press, 2015.

[8] V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. doi: 10.1007/978-1-4612-0537-1. URL `http://dx.doi.org/10.1007/978-1-4612-0537-1`. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

[9] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. URL `http://www.jstatsoft.org/v33/i01/`.

[10] A. Garivier and E. Kaufmann. Optimal best arm identification with fixed confidence. Technical report, HAL, 2016. https://hal.archives-ouvertes.fr/hal-01273838.

[11] Y. Goldberg, R. Song, D. Zeng, and M. R. Kosorok. Comment on "Dynamic treatment regimes: Technical challenges and applications". *Electron. J. Stat.*, 8:1290–1300, 2014.

[12] E. Kaufmann. *Analyse de stratgies baysiennes et frquentistes pour l'allocation squentielle de ressources*. PhD thesis, TELECOM ParisTech, 2014. http://chercheurs.lille.inria.fr/ekaufman/TheseEmilie.pdf.

[13] E. B. Laber, D. J. Lizotte, M. Qian, W. E. Pelham, and S. A. Murphy. Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.*, 8(1):1225–1272, 2014.

[14] E. B. Laber, D. J. Lizotte, M. Qian, W. E. Pelham, and S. A. Murphy. Rejoinder of "Dynamic treatment regimes: Technical challenges and applications". *Electron. J. Stat.*, 8(1):1312–1321, 2014.

[15] A. R. Luedtke and M. J. van der Laan. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference*, 3(1):61–95, 2015.

[16] A. R. Luedtke and M. J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.*, 2015. To appear.

[17] A. R. Luedtke and M. J. van der Laan. Super-learning of an optimal dynamic treatment rule. *International Journal of Biostatistics*, 2016. To appear.

[18] E. Mammen and A. B. Tsybakov. Smooth discrimination analysis. *Ann. Statist.*, 27 (6):1808–1829, 1999. doi: 10.1214/aos/1017939240. URL `http://dx.doi.org/10.1214/aos/1017939240`.

[19] J. Pearl. *Causality: Models, Reasoning and Inference*, volume 29. Cambridge University Press, Cambridge, 2000.

[20] M. Qian and S. A. Murphy. Performance guarantees for individualized treatment rules. *Ann. Statist.*, 39(2):1180–1210, 2011. doi: 10.1214/10-AOS864. URL `http://dx.doi.org/10.1214/10-AOS864`.

[21] J. M. Robins. Optimal structural nested models for optimal sequential decisions. In D. Y. Lin and P. Heagerty, editors, *Proc. Second Seattle Symp. Biostat.*, pages 189–326, 2004.

[22] J. M. Robins and A. Rotnitzky. Discussion of "Dynamic treatment regimes: Technical challenges and applications". *Electron. J. Stat.*, 8(1):1273–1289, 2014.

[23] D. B. Rubin and M. J. van der Laan. Statistical issues and limitations in personalized medicine research with clinical trials. *Int. J. Biostat.*, 8(1), 2012. Article 1.

[24] P. K. Sen and J. M. Singer. *Large sample methods in statistics*. Chapman & Hall, New York, 1993. An introduction with applications.

[25] M. J. van der Laan and S. Rose. *Targeted learning*. Springer Series in Statistics. Springer, New York, 2011. doi: 10.1007/978-1-4419-9782-1. URL `http://dx.doi.org/10.1007/978-1-4419-9782-1`. Causal inference for observational and experimental data.

[26] M. J. van der Laan and D. Rubin. Targeted maximum likelihood learning. *Int. J. Biostat.*, 2:Art. 11, 40, 2006. doi: 10.2202/1557-4679.1043. URL `http://dx.doi.org/10.2202/1557-4679.1043`.

[27] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[28] A. W. van der Vaart and J. A. Wellner. *Weak Convergence*. Springer, 1996.

[29] B. Zhang, A. Tsiatis, M. Davidian, M. Zhang, and E. Laber. A robust method for estimating optimal treatment regimes. *Biometrics*, 68:1010–1018, 2012.

[30] B. Zhang, A. Tsiatis, M. Davidian, M. Zhang, and E. Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 68(1):103–114, 2012.

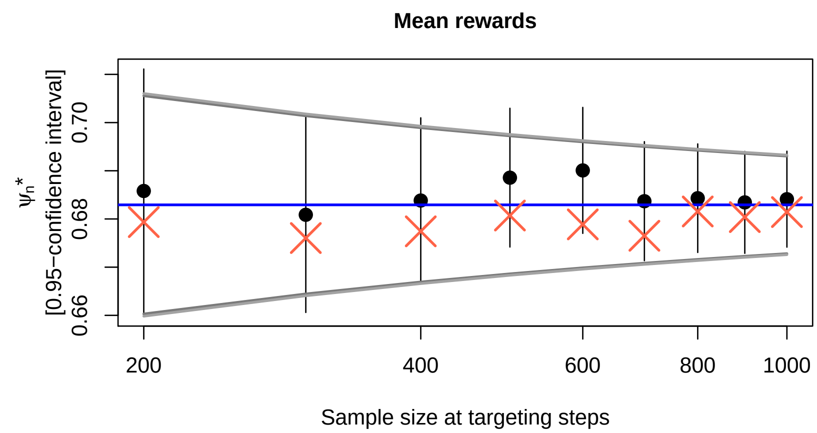
[31] Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.*, 107(499): 1106–1118, 2012. doi: 10.1080/01621459.2012.695674. URL http://dx.doi.org/10.1080/01621459.2012.695674.

[32] Y. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.*, 110(510): 583–598, 2015. doi: 10.1080/01621459.2014.937488. URL http://dx.doi.org/10.1080/01621459.2014.937488.

[33] W. Zheng, A. Chambaz, and M. J. van der Laan. Drawing valid targeted inference when covariate-adjusted response-adaptive rct meets data-adaptive loss-based estimation, with an application to the LASSO. Technical Report 339, U.C. Berkeley Division of Biostatistics Working Paper Series, 2015. URL http://biostats.bepress.com/ucbbiostat/paper339.

parameter		intermediate sample sizes n								
		200	300	400	500	600	700	800	900	1000
$\psi_{r_n,0}$	a	928	939	933	945	948	955	951	946	943
	b	0.0015	0.0671	0.0106	0.2529	0.4058	0.7853	0.5780	0.3002	0.1721
	c	0.0658	0.4657	0.1920	0.7653	0.8725	0.9837	0.9407	0.8053	0.6739
	d	0.6754	0.6739	0.6735	0.6797	0.6796	0.6798	0.6799	0.6800	0.6800
	e	0.2039	0.1980	0.1947	0.1849	0.1841	0.1856	0.1825	0.1826	0.1835
ψ_0	a	932	927	916	949	952	941	950	944	946
	b	0.0074	0.0010	0.0001	0.4625	0.6344	0.1106	0.5203	0.2101	0.3002
	c	0.1588	0.0512	0.0014	0.8994	0.9558	0.5718	0.9221	0.7213	0.8053
	d	0.6827	0.6827	0.6827	0.6827	0.6827	0.6827	0.6827	0.6827	0.6827
	e	0.2039	0.1980	0.1947	0.1849	0.1841	0.1856	0.1825	0.1826	0.1835
$\mathcal{E}_n$	a	967	981	983	970	979	973	968	965	967
	b	0.9963	1	1	0.9993	1	0.9999	0.9978	0.9907	0.9963
	c	1	1	1	1	1	1	1	0.9999	1
	d	-0.1411	-0.0936	-0.0705	-0.0629	-0.0529	-0.0460	-0.0406	-0.0365	-0.0332
	e	1.2984	1.3011	1.3258	1.2650	1.2848	1.2976	1.3012	1.3104	1.3567
$\mathcal{C}_n$	a	989	996	998	995	999	999	999	999	1000
	b	1	1	1	1	1	1	1	1	1
	c	1	1	1	1	1	1	1	1	1
	d	-0.1417	-0.0942	-0.0711	-0.0634	-0.0533	-0.0463	-0.0410	-0.0368	-0.0335
	e	1.2925	1.2902	1.3096	1.2525	1.2707	1.2833	1.2819	1.2915	1.3287

Table 1: Description of the results across the $N = 1000$ repeated simulations. Rows ^{a}: empirical coverages. Rows ^{b} and ^{c}: p-values of the corresponding binomial tests of coverage at least 95% and at least 94% (see Section 6.2). Rows ^{d}: mean values of the possibly data-adaptive parameters. Rows ^{e}: from top to bottom, mean values of Σ_n (twice), mean values of $|\mathcal{E}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{E}_n|$, mean values of $|\mathcal{C}_n - (n^{-1}\sum_{i=1}^n Y_i - \psi_n^* + \xi_\alpha\sqrt{\Sigma_n^{\mathcal{E}}/n})|/|\mathcal{C}_n|$.



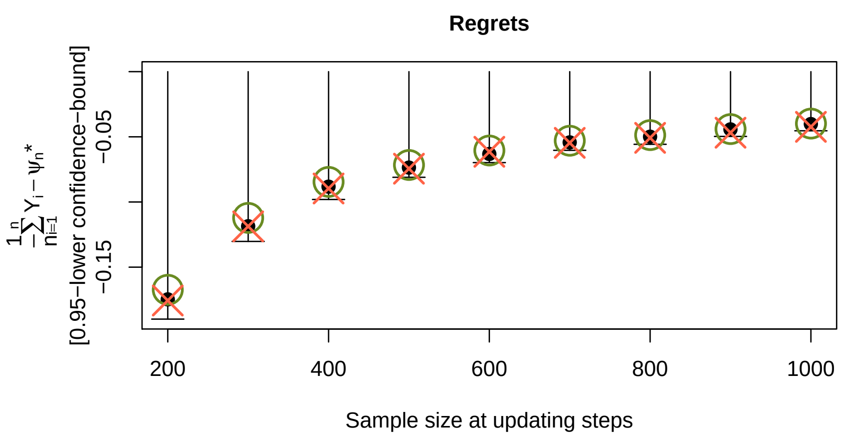


Figure 1: **Illustrating the data-adaptive inference of the optimal treatment rule, its mean reward and the related pseudo-regrets** (see also Figure 2). *Top plot.* The blue horizontal line represents the value of the mean reward under the optimal treatment rule, ψ_0. The grey curves represent the mapping $n \mapsto \psi_0 \pm \xi_{97.5\%}\sigma_0/\sqrt{n}$, where $\sigma_0 = 0.1634$ is the square root of $\mathrm{Var}_{P_{Q_0,r_0}} D(Q_0, r_0)(O)$; thus, at a given sample size n, the length of the vertical segment joining the two curves equals the length of a confidence interval based on a regular, asymptotically efficient estimator of ψ_0. The pink crosses represent the successive values of the data-adaptive parameters $\psi_{r_n,0}$. The black dots represent the successive values of ψ_n^*, and the vertical segments centered at them represent the successive 95%-confidence intervals for $\psi_{r_n,0}$ and, under additional assumptions, for ψ_0 as well. *Bottom plot.* The pink crosses and green circles represent the successive values of the empirical and counterfactual cumulative pseudo-regrets $\mathcal{E}_n$ and $\mathcal{C}_n$. The black dots represent the successive values of $n^{-1}\sum_{i=1}^n Y_i - \psi_n^*$, and the vertical segments represent the successive 95%-lower confidence bounds on $\mathcal{E}_n$ and $\mathcal{C}_n$.

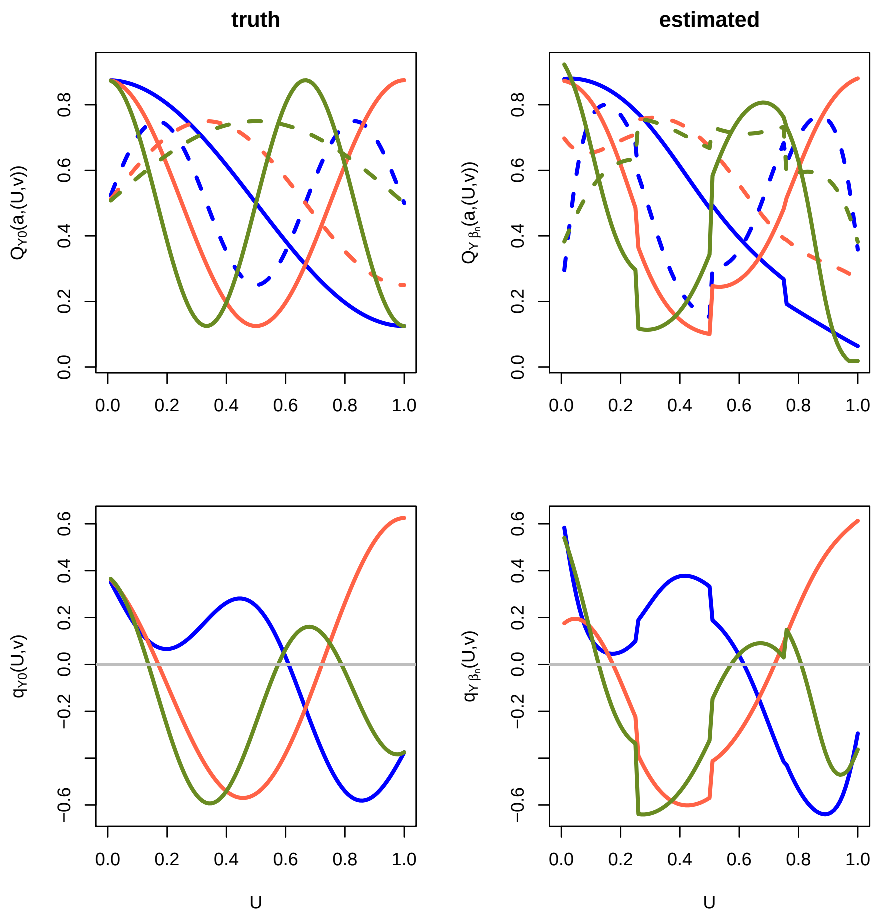


Figure 2: **Illustrating the data-adaptive inference of the optimal treatment rule, its mean reward and the related pseudo-regrets through the representation of the conditional mean $Q_{Y,0}$, blip function $q_{Y,0}$ and their estimators** (see also Figure 1). *Top left plot.* The solid curves represent $U \mapsto Q_{Y,0}(1,(U,v))$ for $v=1$ (in blue, minimum reached at $U=1$), $v=2$ (in pink, minimum reached at $U=1/2$) and $v=3$ (in green, minimum reached at $U=1/3$). The dashed curves represent $U \mapsto Q_{Y,0}(0,(U,v))$ for $v=1$ (in blue, maximum reached at $U=1/6$), $v=2$ (in pink, maximum reached at $U=1/3$) and $v=3$ (in green, minimum reached at $U=1/2$). *Bottom left plot.* The curves represent $U \mapsto q_{Y,0}(U,v)$ for $v=1$ (in blue, minimum reached close to $1/9$), $v=2$ (in pink, minimum reached close to $1/2$) and $v=3$ (in green, minimum reached close to $1/3$). *Right plots.* Counterparts to the left plots, where $Q_{Y,0}$ and $q_{Y,0}$ are replaced with Q_{Y,β_n} and q_{Y,β_n} for $n=1000$, the final sample size.